

MMV
IMM
220

CURTIS

MATRIX INVERSION ...



NEW YORK UNIVERSITY
INSTITUTE OF MATHEMATICAL SCIENCES
LIBRARY
25 Waverly Place, New York 3, N. Y.

IMM-NYU 220
August 1955

MATRIX INVERSION AND THE SOLUTION OF LINEAR EQUATIONS

by

J. H. Curtiss

25

REPRODUCTION IN WHOLE OR IN PART
IS PERMITTED FOR UNLIM-
ITED USE
OF THE UNITED STATES GOVERNMENT.

Contracts Nos. DA-30-069-ORD-1257

and

N6ori-201, Task Order No. 1

Technical Report No. V

This report represents results obtained at the
Institute of Mathematical Sciences, New York University,
under the sponsorship of the Army Office of Ordnance
Research and the Office of Naval Research.

New York, 1955

Chapter 1.

MATRICES, VECTORS, AND DETERMINANTS

1.1 The fundamental existence theorem.

Theorem 1.1 . Consider the following system of n simultaneous linear algebraic equations* :

$$(1.1.1) \quad \begin{cases} a_{11}x^1 + a_{12}x^2 + \dots + a_{1n}x^n = y^1 \\ a_{21}x^1 + a_{22}x^2 + \dots + a_{2n}x^n = y^2 \\ \vdots \\ a_{n1}x^1 + a_{n2}x^2 + \dots + a_{nn}x^n = y^n \end{cases}$$

in which the symbols a_{ij} , $i, j = 1, \dots, n$ and y^i , $i = 1, \dots, n$, represent real or complex numbers, and the symbols x^i , $i = 1, \dots, n$, represent unknowns.

With a fixed set of numbers a_{ij} , either

(1) There is one and only one solution $(x^1, x^2, \dots, x^n)$ for each and every assignment of the numbers y^i , or else

(2) With $y^1 = y^2 = \dots = y^n = 0$, there are k , $1 \leq k < n$, solutions $(x^1, x^2, \dots, x^n)$ with $|x^1| + |x^2| + \dots + |x^n| \neq 0$, which are all different in the sense that they are linearly dependent**. In this case, the system of simultaneous linear equations

* The superscripts are indices, not exponents.

** This term is explained in Section 1.5 below.

$$(1.1.2) \quad \left\{ \begin{array}{l} a_{11}t_1 + a_{21}t_2 + \dots + a_{n1}t_n = 0 \\ a_{12}t_1 + a_{22}t_2 + \dots + a_{n2}t_n = 0 \\ \cdot \\ \cdot \\ \cdot \\ a_{1n}t_1 + a_{2n}t_2 + \dots + a_{nn}t_n = 0 \end{array} \right.$$

also has k linearly independent solutions $(t^1, t^2, \dots, t^k)$, with $|t^1| + |t^2| + \dots + |t^k| \neq 0$; and a necessary and sufficient condition that (1.1.1) have solutions for a set of numbers t^i not all zero is that for every one of the k linearly independent solutions of (1.1.2), the relation $y^1t^1 + y^2t^2 + \dots + y^nt^n = 0$ is satisfied.

There are a number of other ways in which the basic facts concerning the existence of solutions to systems of linear algebraic equations can be stated. The formulation above has the advantage of telling the story with a minimum amount of technical terminology. Also, it is of interest to mathematical analysts because the existence theorems for integral equations can be given a parallel formulation.

The proof of the theorem will be constructed in the course of the chapter. At the same time a number of other fundamental results will be derived, which are essential for an understanding of the theory of linear computations.

1.2* Matrices and their algebra . A finite $n \times m$ matrix is an

* The meaning of the superscript c preceding the section number is that all statements and results in the section, when not specifically restricted, are valid for matrices with complex elements as well as for matrices with real elements.

ordered set of mn numbers a^{ij} , $i = 1, \dots, n$, $j = 1, \dots, m$. They are usually displayed in the form of a rectangular table as follows:

$$(1.2.1) \quad \begin{bmatrix} a^{11} & a^{12} & \dots & a^{1m} \\ a^{21} & a^{22} & & a^{2m} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ a^{n1} & a^{n2} & \dots & a^{nm} \end{bmatrix}$$

The first superscript of a^{ij} then represents the row and the second superscript the column in which this element of the matrix is located.

If $n = m$, the matrix is called square. In this case, the elements a^{ii} , $i = 1, \dots, n$, form what is called the principal diagonal of the matrix.

Various symbols of enclosure other than the square brackets used in (1.2.1) are in common use; for example,

$$\| \quad \|, \quad (\quad) .$$

The briefer notations $[a^{ij}]$, (a^{ij}) , . . . are also often used for (1.2.1). Moreover, it is common to use a single letter, such as A , for a matrix. The enclosure symbols which we shall ordinarily employ in this book will be square brackets, unless a matrix has only one row or one column. Such a matrix is called a row matrix, or a column matrix, as the case may be. In the latter case, if the elements are $x^1, x^2, \dots, x^n$, we shall use the notation $(x^1, x^2, \dots, x^n)$ to represent the

matrix. If a single letter is used to denote such a matrix, it will usually be a lower case Greek letter, such as ξ .

The elements of the matrices dealt with in this book may always be thought of as real numbers. However, a number of the statements and results pertain also to matrices with complex elements, or for that matter to matrices with elements lying in even more general fields. If the elements of a matrix are to be considered as complex numbers, and if it is desired to emphasize this fact, we shall call the matrix a complex matrix. When complex matrices are in view, a matrix with real elements will be distinguished by calling it a real matrix.

We shall use the notation $\bar{z}$ for the complex conjugate of a complex number z . (That is, if $z = x + iy$, with x and y real and $i^2 = -1$, then $\bar{z} = x - iy$.) The modulus or absolute value of z will be denoted by $|z|$. (That is, $|z| = (x^2 + y^2)^{1/2}$.) The matrix obtained by replacing each element a^{ij} in the complex matrix in $A = [a^{ij}]$ by $\bar{a}^{ij}$ will be denoted by $\bar{A}$.

If all the elements of a matrix A are zero, we shall call the matrix a zero matrix and use the symbol O for it.

A square matrix with only zeros off the principal diagonal is called a diagonal matrix. Thus a zero matrix is a special kind of diagonal matrix. Another special kind of diagonal matrix is a ^{diagonal} matrix in which the principal diagonal is composed entirely of ones. This is called an identity ma-

trix, or unit matrix, and will be denoted by I , or by I_n if it is desired to emphasize the number of rows and columns.

If all the elements of a square matrix above the principal diagonal are zeros, the matrix is called a lower triangular matrix. There is a similar definition for an upper triangular matrix. If the elements on the principal diagonal of a lower triangular matrix are all ones, then it will be called a lower unit triangle. An upper unit triangle is similarly defined.

Two $n \times m$ matrices A and B are equal if and only if each and every element of A equals the corresponding element of B . Thus the matrix (1.2.1) would be changed into a different matrix if two unequal elements were to be interchanged in the rectangular table. If A is equal to B , we naturally write $A = B$.

The sum of two matrices is defined only when they each have the same number of rows and columns. Then the sum $A + B$ of $A = [a^{ij}]$ and $B = [b^{ij}]$, is defined to be the matrix $[a^{ij} + b^{ij}]$. It is obvious that

$$A + (B + C) = (A + B) + C$$

$$A + B = B + A$$

$$A + O = A$$

The difference of two matrices A and B having the same number of rows and columns is defined by

$$A - B = [a^{ij} - b^{ij}].$$

The product of two matrices is defined only on the as-

sumption that the number of columns of the first factor in the product equals the number of rows of the second factor. In that case the matrices are called conformable. Let $A = [a^{ij}]$ be $n \times m$ and $B = [b^{ij}]$ be $m \times p$. The product AB is then defined to be the $n \times p$ matrix $[c^{ij}]$, where

$$c^{ij} = a^{i1} b^{1j} + a^{i2} b^{2j} + \dots + a^{im} b^{mj}.$$

In words, the element in the i, j -th position of c^{ij} is the sum of the products of the elements in the i -th row of A and j -th column of B .

It will be noticed that the product of two square matrices of the same order is again a square matrix of the same order; the product of an $n \times m$ matrix into a matrix consisting of only one column containing m elements is a matrix consisting of only one column of n elements, and the product of a matrix consisting of only one row of n elements into an $n \times p$ matrix is a matrix consisting of a row of p elements. Other facts of this type will be found in Exercise 3 at the end of this section.

If ξ denotes a matrix consisting of just one column, whose elements are $x^1, x^2, \dots, x^m$, and η is another column matrix, but with the elements $y^1, y^2, \dots, y^n$, and A denotes the matrix of the coefficients of the system of simultaneous equations (1.1.1), then it is readily seen that we can rewrite the system in the compact form

$$(1.2.2) \quad A\xi = \eta.$$

It is easy to verify by direct calculation that

$$A(BC) = (AB)C$$

$$(A + B)C = AC + BC$$

$$C(A + B) = CA + CB \quad .$$

In the product AB , A is said to be postmultiplied by B , and B is said to be premultiplied by A .

In general, $AB \neq BA$; matrix multiplication is ordinarily not commutative. If it should happen that $AB = BA$, then A and B are said to commute, or to be commutative.

In the special case of a diagonal matrix C whose principal diagonal elements are all equal to the same number c , it is easily checked that multiplication with any other square matrix $A = [a^{ij}]$ is commutative. In fact, $AC = CA = [ca^{ij}]$. This suggests that we might as well denote CA by cA , and define thereby the process of multiplying a matrix by an ordinary number. The process is called scalar multiplication; the matrix C is called a scalar matrix; and when a number c is to be associated with matrices in scalar multiplications, it is often called a scalar.

Evidently I is a special kind of scalar matrix. It has the property that $AI = IA = A$. Thus it plays the same role in the algebra of matrices that the number one does in the algebra of real numbers.

Although the algebra of matrices as defined above has many similarities with that of ordinary numbers, the reader is warned to observe two particularly important differences. The first is that multiplication does not follow the commutative law. The second is that the product of two matrices, neither of which is

a zero matrix, can be a zero matrix. We return to this matter in Exercise 2 at the end of the section.

If the matrix $B = [b^{ij}]$ is obtained from the $n \times m$ matrix $A = [a^{ij}]$ by using the rows of A (in order) for the columns in B — that is, if $b^{ij} = a^{ji}$, $i = 1, \dots, m$, $j = 1, \dots, n$ — then B is called the transpose of A . We shall use the notation A' for the transpose of A .

If A is square and $A' = A$, then A is called symmetric. If A is square and complex, and if $\bar{A}' = A$, then A is called hermitian. The principal diagonal elements of a hermitian matrix are necessarily real.

If A is an $n \times n$ matrix, and if B is another $n \times n$ matrix such that $BA = I$ and $AB = I$, then B is called the inverse (or reciprocal) of A , and we write $B = A^{-1}$. Clearly A is also the inverse of B , according to this definition. The justification for calling B "the" inverse is given in the following theorem.

Theorem 1.2 . If A is an $n \times n$ matrix, and if B and C are two $n \times n$ matrices such that $BA = I$ and $AC = I$, then $B = C$.

The proof consists merely in postmultiplying the equation $BA = I$ by C and premultiplying the equation $AC = I$ by B . We obtain $BAC = C$, $BAC = B$; so $B = C$.

The theorem clearly implies the uniqueness of an inverse B of A . For any other inverse B_0 would satisfy the equation $AB_0 = I$. Since B itself satisfies the equation $BA = I$, it follows from the theorem that $B = B_0$.

Not all square matrices have inverses. For example, the matrix

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$$

cannot have an inverse, because for any matrix B the second row of the product AB must consist of two zeros instead of a zero and a one.

If a matrix does have an inverse, we call it non-singular; otherwise we call it singular.

It will turn out in Section 1.3, Exercise 13, that our definition of inverse is somewhat redundant. The fact is that it suffices to define the inverse of A as a matrix B such that $BA = I$; then automatically it follows that $AB = I$. Similarly, the inverse can be defined as a matrix B such that $AB = I$; then automatically $BA = I$.

The title of this book indicates that one of the chief problems to be discussed will be that of calculating the matrix A^{-1} , given the matrix A . The relationship of this problem to the solution of linear simultaneous equations is easily developed. Write the system (1.1.1) in the form (1.2.2):

$$A\xi = \eta.$$

Let us suppose that A is non-singular. Then obviously the column matrix $\xi = A^{-1}\eta$ is a solution, for substituting it into the equations, we obtain $A(A^{-1}\eta) = I\eta = \eta$. Moreover if ξ_0 is any other solution of (1.1.1), then premultiplying the relation $A\xi_0 = \eta$ by A^{-1} establishes that $\xi_0 = A^{-1}\eta$. So

in the non-singular case, a solution not only exists; it is also unique. This is alternative (1) of Theorem 1.1 .

On the other hand, if A is non-singular, the construction of a matrix B such that $AB = I$ can be carried out by solving n systems of n simultaneous linear equations, one system for each column of B .

We shall discuss the relationship of the two central problems of this book in further detail in Section 1.6 below.

There are many important and useful results in elementary matrix theory which are quite immediate consequences of the definitions so far presented. We shall now list a number of them in the form of exercises, which the reader who is unfamiliar with matrices should work out for himself.

Exercises.

1. Let the variables $y^1, y^2, \dots, y^n$ and $x^1, x^2, \dots, x^n$ be related by the linear transformation

$$y^i = \sum_{k=1}^n b^{ik} x^k, \quad i = 1, \dots, n.$$

Let $z^1, z^2, \dots, z^n$ be related to $y^1, y^2, \dots, y^n$ by

$$z^i = \sum_{k=1}^n a^{ik} y^k, \quad i = 1, \dots, n.$$

By substitution, show that the coefficients in the linear transformation which gives the z 's directly in terms of the x 's, are merely the elements of AB . (This is the motivation for the definition of matrix multiplication.)

2. Show by an example that matrices $A \neq 0$ and $B \neq 0$ exist such

that $AB = O$.

3a. Show that the product of two diagonal matrices D_1 and D_2 is a diagonal matrix, each of whose elements is the product of the corresponding elements of D_1 and D_2 .

b. Show that the product of two lower (upper) triangular matrices is a lower (upper) triangular matrix, and that if the triangular matrices are unit triangles, so is their product.

4. Show by an example that A and A' are not necessarily commutative.

5. Let D be an $n \times n$ diagonal matrix and let A be any $n \times n$ matrix. Write out the products DA and AD and thereby show that they will not in general be equal. Show that if however A is a diagonal matrix, then $DA = AD$.

6. Show that $(AB)' = B'A'$, where A and B are any two conformable rectangular real or complex matrices. (This is the reversal rule for transposes. Hint: Simply write out the (i, j) -th term of the matrix on the left side of the equation, and of the matrix on the right side.)

7. Use 6. to prove that $(A_1 A_2 \dots A_k)' = A_k' A_{k-1}' \dots A_1'$.

8. Let A be an arbitrary $n \times m$ matrix. Show that $B_1 = AA'$ and $B_2 = A'A$ are always symmetric matrices, whether or not A is symmetric. (Hint: Use 6. and the fact that $(A')' = A$.)

9. Let A and B be conformable complex matrices. Prove that $(\overline{AB}) = \overline{A} \overline{B}$.

10. Let A be an arbitrary rectangular complex matrix. Show that $B_1 = A\overline{A}'$ and $B_2 = \overline{A}'A$ are hermitian.

11. Prove that if A and B are symmetric, so is $ABA \dots BA$. Show by example (using matrices of order 2×2) that if A and B are both symmetric, this does not necessarily imply that AB is.
12. Prove that if A is symmetric, then B^tAB is also symmetric for any square matrix B . (Note: If for some matrix B , $C \equiv B^tAB$, then C is said to be congruent to A .)
13. Prove that every square matrix A is uniquely expressible as $A \equiv H + K$, where H is symmetric and $K \equiv -K^t$. (Note: A matrix having this latter property is called skew-symmetric. The implication in a skew-symmetric matrix is that $a_{ii} = 0$, $i = 1, \dots, n$. Hint: Is $A + A^t$ symmetric? Is $A - A^t$ skew-symmetric?)

Note: All matrices in exercises 14. - 22. are $n \times n$ unless a statement is made to the contrary.

14. Show that if D is a diagonal matrix with principal diagonal elements $d^1 \neq 0$, $d^2 \neq 0$, $\dots$, $d^n \neq 0$, then D is non-singular, and D^{-1} is a diagonal matrix with principal diagonal elements $1/d^1$, $1/d^2$, $\dots$, $1/d^n$.
15. Prove that if A is non-singular, then A^t is non-singular and $(A^t)^{-1} \equiv (A^{-1})^t$. (Hint: Apply 14. and the reversal rule for transposes to $A^{-1}A \equiv I$ and $AA^{-1} \equiv I$.)
16. If A and B are both non-singular, prove that $(AB)^{-1}$ exists and equals $B^{-1}A^{-1}$. (The reversal rule for inverses. Hint: Multiply $B^{-1}A^{-1}$ by AB .)
17. Use 16. to show that if $A_1, A_2, \dots, A_k$ are non-singular, then $(A_1A_2 \dots A_k)^{-1}$ exists and equals $A_k^{-1}A_{k-1}^{-1} \dots A_1^{-1}$.
18. Let A be an $n \times m$ matrix, let ξ be an $m \times 1$ matrix, let η be an $n \times 1$ matrix, and let B be a non-singular $n \times n$ matrix.

Prove that ξ is a solution of the system $A\xi = \eta$ if and only if $B\xi = B\eta$ is a solution of $BA\xi = B\eta$. (Note: The premultiplication of a system $A\xi = \eta$ by a non-singular matrix B is called preconditioning the system. By taking $m = n$, $B = A'$, the exercise shows that a system $A\xi = \eta$ with a non-singular square matrix can always be replaced by an equivalent system with a symmetric matrix.

19. Prove that if A and B are commutative and A is non-singular, then A^{-1} commutes with B .

20. Prove that the inverse of a lower (upper) triangular matrix is a lower (upper) triangular matrix, that the inverse of a lower (upper) unit triangle is a lower (upper) unit triangle, that the inverse of a symmetric matrix is also symmetric, and that the inverse of a skew-symmetric matrix of even order is also skew-symmetric. (Note: A skew-symmetric matrix of odd order is necessarily singular, this will be brought out later in the chapter.)

21. We define A^N to mean $\overbrace{A \cdot A \cdots A}^N$ and A^{-N} as $\overbrace{A^{-1} \cdot A^{-1} \cdots A^{-1}}^N$, where N is any positive integer; further we define A^0 to be I . From these definitions and the laws obeyed by matrix multiplication, establish that the laws of exponents $A^N A^M = A^{N+M}$, $(A^N)^M = A^{NM}$, hold for arbitrary integers N and M .

22. Let B be any non-singular $n \times n$ matrix, and let A be an arbitrary $n \times n$ matrix. Suppose further that all inverses indicated in the list below exist. Prove that

$$(a) \quad (B^{-1}AB)^N = B^{-1}A^NB$$

$$(b) \quad (I - B^{-1}AB)^{-1} = B^{-1}(I - A)^{-1}B$$

$$(c) \quad B(I - AB)^{-1} = (I - BA)^{-1}B$$

$$(d) \quad B(I - AB)^{-1}B^{-1} = (I - BA)^{-1}$$

$$(e) \quad B^{-1}(I - BA)^{-1}B = (I - AB)^{-1}$$

Note: If for some matrix B , $C = B^{-1}AB$, then C is said to be similar to A , or to be obtained from A by a similarity transformation.

23. If $p(x) = c_0x^N + c_1x^{N-1} + \dots + c_N$, where all letters here stand for scalars, we define $p(A)$ to mean $c_0A^N + c_1A^{N-1} + \dots + c_N$, where A is an arbitrary square matrix. Prove that $B^{-1}p(A)B = p(B^{-1}AB)$.

24. It is the purpose of this exercise to outline a proof of the following result: Suppose that the two conformable matrices A and B are partitioned into submatrices or blocks

A_{ij} and B_{ij} as follows:

$$A = \begin{bmatrix} A_{11} & | & A_{12} & | & \dots & | & A_{1k} \\ \hline - & - & - & - & - & - & - \\ A_{21} & | & A_{22} & | & \dots & | & A_{2k} \\ \hline - & - & - & - & - & - & - \\ \vdots & | & & | & & | & \\ \vdots & | & & | & & | & \\ \hline A_{h1} & | & A_{h2} & | & \dots & | & A_{hk} \\ & | & & | & & | & \end{bmatrix}$$

$$B = \begin{bmatrix} B_{11} & B_{12} & \dots & B_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ B_{k1} & B_{k2} & \dots & B_{kk} \end{bmatrix}$$

Suppose further that the rule governing the partitioning is that whenever a dashed line is inserted between two columns of A , say the q -th and $(q+1)$ -th, a dashed line is inserted between the q -th and $(q+1)$ -th rows of B , and the number of dashed lines between the columns of A and rows of B is the same. Then whatever the horizontal partitioning of A and the vertical partitioning of B , the product AB can be expressed as a partitioned matrix C , partitioned according to the horizontal partitioning of A and the vertical partitioning of B ; and the (i, j) -th submatrix of C is $C_{ij} = A_{i1}B_{1j} + A_{i2}B_{2j} + \dots + A_{ik}B_{kj}$. The steps in the proof are as follows:

(a) Let A be $n \times m$ and B be $m \times p$. Partition A into two blocks A_1 and A_2 by a dashed line between any two rows. Verify that

$$AB = \begin{bmatrix} A_1 B \\ A_2 B \end{bmatrix}$$

(b) Partition A by vertical dashed lines only, into submatrices $A_1, A_2, \dots, A_k$. Partition B by correspondingly placed horizontal dashed lines into submatrices $B_1, \dots, B_k$.

Prove that $AB = A_1B_1 + A_2B_2 + \dots + A_kB_k$.

(Hint: Just examine a typical element of the product, written out at length, and put parentheses into it strategically.)

(c) Combine Exercise (a) and (b) to get the general result.

25. Let

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

where each submatrix is $n \times n$, and let

$$B = \begin{bmatrix} 0 & I_n \\ I_n & 0 \end{bmatrix}$$

Show that

$$BA = \begin{bmatrix} A_{21} & A_{22} \\ A_{11} & A_{12} \end{bmatrix}$$

and

$$AB = \begin{bmatrix} A_{12} & A_{11} \\ A_{22} & A_{21} \end{bmatrix}$$

26. Let A be an $n \times n$ matrix and let η be a $n \times 1$ matrix.

Let B be the following $(n+1) \times (n+1)$ matrix:

$$B = \begin{bmatrix} A & -\eta \\ 0 & 1 \end{bmatrix}$$

(Here 0 denotes the $1 \times n$ zero matrix.) Show by multiplication that

$$B^{-1} = \begin{bmatrix} A^{-1} & A^{-1}\eta \\ 0 & 1 \end{bmatrix}$$

(So the upper blocks of B^{-1} contain the solution to the problem of finding the inverse of A and also to that of solving the system $Ax = \eta$. This exercise is generalized in Chapter 3.)

1.3 Elementary operations and the reduction of a matrix to triangular form. We identify three kinds of elementary row operations on matrices:

Type (i). Interchange of any two rows.

Type (ii). Multiplication of any row by a non-zero scalar.

Type (iii). Addition of a scalar multiple of any row to any other row.

Let A be an arbitrary $n \times m$ matrix. Let U_{ij} denote an $n \times n$ matrix in which the elements are all zeros, except for the (i, j) -th, which is unity. The reader can easily verify the truth of the following theorems by performing the implied matrix multiplications.

Theorem 1.3.

(a) Premultiplying A by $I = U_{ii} = U_{jj} \oplus U_{ij} \oplus U_{ji}$ will interchange the i-th and j-th rows of A. (The Type (i) elementary row operation.)

(b) Premultiplying A by $I \oplus cU_{ii}$ will multiply the i-th row by $1 \oplus c$. (The Type (ii) elementary row operation if $c \neq -1$.)

(c) Premultiplying A by $I \oplus cU_{ij}$, $i \neq j$, will replace the i-th row of A by the sum of the i-th row and c multiplied into the j-th row. (The Type (iii) elementary row operation.)

The matrix premultipliers which appear in Theorem 1.3 (with $c \neq -1$ in (b)) are called elementary matrices. The elementary matrix $I = U_{ii} = U_{jj} \oplus U_{ij} \oplus U_{ji}$ which appears in (a) is merely the unit matrix with the 1 in the i-th row moved over to the j-th column, and the 1 in the j-th row moved over to the i-th column. The elementary matrix $I \oplus cU_{ii}$ which appears in (b) is a diagonal matrix, and the elementary matrix $I \oplus cU_{ij}$, $i \neq j$, which appears in (c) is a unit triangle.

Theorem 1.4 . Each elementary matrix has an inverse, which itself is an elementary matrix. The inverse corresponds to the elementary row operation which "undoes" the effect of premultiplying by the original elementary matrix. Thus the inverse corresponds to an elementary row operation of the same type as that to which the original elementary row operation corresponds.

The truth of Theorem 1.4 is intuitively apparent. The explicit construction of the inverses is an easy matter. (Exercise 1. at the end of this section.)

There are three analogous elementary column operations, which can be performed by postmultiplying a matrix A by (con-

formable) elementary matrices. The elementary column operations will receive less emphasis in this book than the row operations. This is because of our preoccupation with the problem $A\xi = \eta$, in which column operations and postmultiplications of A by other matrices are less convenient to deal with than row operations and pre-multiplications of A .

Each particular elementary matrix can be constructed by performing the corresponding elementary operation on the unit matrix I .

A matrix B is called row-equivalent to a matrix A if it can be obtained from A by a finite sequence of elementary row operations. A matrix B is called column-equivalent to a matrix A if it can be obtained from A by a finite sequence of elementary column operations. Since each elementary operation is paired with another elementary operation which undoes its effect, if B is row- or column- equivalent to A , then A is row- or column- equivalent to B .

We now come to a result which is perhaps the most important one in the entire theory of linear computations. In fact, some persons hold the view that once this result has been established, the computational problems which form the subject of this book are completely obviated, and nothing more needs to be said about them!

Let A be an arbitrary $n \times m$ matrix. In any non-zero row of A , we call the first non-zero element the leading element of that row. (A zero row has no leading element.) We shall say that A is semi-row-reduced if every column containing a leading element of some row has all its other elements zero. A sample

semi-row-reduced matrix is as follows:

$$\begin{bmatrix} 0 & 0 & 0 & b^1 & 0 & a^{15} & a^{16} \\ 0 & b^2 & a^{23} & 0 & c & a^{25} & a^{26} \\ 0 & 0 & 0 & 0 & b^3 & a^{35} & a^{36} \\ 0 & 0 & 0 & 0 & 0 & c & 0 \end{bmatrix}$$

The letters b^1 , b^2 , b^3 stand for non-zero elements.

If the leading elements of a semi-row-reduced matrix are all ones, we call the matrix row-reduced^{*}. Finally, if each zero row of a matrix comes below all the non-zero rows, and if $j_1 < j_2 < \dots < j_r$, where $j_1, j_2, \dots, j_r$ are the column indices of the leading elements, then we call the matrix an echelon matrix.

The sample matrix given above can be brought into echelon form by interchanging the first and second rows.

The important result advertised above is as follows:

Theorem 1.5 .

- (a) Any matrix A is a row-equivalent to a semi-row-reduced matrix under elementary row operations of Type (iii) alone.
- (b) Any semi-row-reduced matrix is equivalent to a semi-row-reduced echelon matrix under row operations of Type (i) alone.
- (c) Any semi-row-reduced matrix is equivalent to a row-reduced matrix under row operations of Type (ii) alone.

Parts (b) and (c) of the theorem are obvious from the definitions. The heart of the theorem lies in part (a), which we now prove.

* We are generally following the terminology of Birkhoff-MacLane (1953), pp. 172-175.

Let i_1 be the index of the first non-zero row of A . Let its leading element be $a_{i_1 j_1}$. Multiply this row by $a_{i_1 j_1} / a_{i_1 j_1}$ and subtract from the i -th row for each $i > i_1$. (Some of the multipliers $a_{i j_1} / a_{i_1 j_1}$ may be zero.) This reduces each element in the j_1 -th column to zero. Let $A_2 = [a_2^{ij}]$ denote the matrix into which A has now been transformed.

Now let i_2 be the index of the first non-zero row of A_2 after the i_1 -th. Let the leading element of the i_2 -th row be $a_{i_2 j_2}$. Multiply this row by $a_{i_2 j_2} / a_{i_2 j_2}$ and subtract from the i -th row for each $i \neq i_2$. This reduces each element in the j_2 -th column to zero, and it does not affect the j_1 -th column (which contains the leading element of the i_1 -th row) because the i_2 -th row has a zero in that column. Let $A_3 = [a_3^{ij}]$ denote the matrix into which A_2 has now been transformed. Note that the first i_2 rows of A_3 , together with all subsequent zero rows, form a semi-row-reduced submatrix of A_3 .

Now let i_3 be the index of the first non-zero row of A_3 after the i_2 -th. Let the leading element of the i_3 -th row be $a_{i_3 j_3}$. Multiply this row by $a_{i_3 j_3} / a_{i_3 j_3}$ and subtract from the i -th row for each $i \neq i_3$. This reduces each element in the j_3 -th column to zero, and it does not affect either the j_1 -th or j_2 -th columns, because the i_3 -th row has zeros in each of these two positions. Note that the first i_3 rows of the new matrix, together with all subsequent zero rows, form a semi-row-reduced submatrix.

The same construction is applied over and over to the successive transformations of A . After the k -th step, the first i_k rows form, together with all zero rows lying below them, a semi-row-reduced submatrix. Since i_k increases at each step,

and since A has only a finite number of rows, it follows that in a finite number of steps the entire matrix A assumes a semi-row-reduced form. Note that operations only of Type (iii) are being used.

The proof of the theorem is complete.

The theorem shows that given any $n \times m$ matrix A , there exists an $n \times n$ matrix $E = E_2 E_1 E_3$ in which E_2 is a diagonal matrix, E_1 is a product of elementary matrices corresponding to elementary row operation of Type (i), and E_3 is a product of elementary matrices corresponding to elementary row operations of Type (iii), which has the property that $E_2 E_1 E_3 A$ is a row-reduced echelon matrix.

The leading elements $a_{i_1 j_1}, a_{i_2 j_2}, \dots, a_{i_r j_r}$ of the semi-row-reduced matrix constructed for part (a) of the theorem by the elementary row operations of Type (iii) are called pivots. It should be noted that the leading elements in the rows $i_1, i_2, \dots, i_r$, of the diagonal matrix E_2 are reciprocals of the pivots, and the leading elements in the other rows of E_2 are ones.

Finally, we observe that E is non-singular. This follows from Exercise 17 of Section 1.2 and the fact that E is a product of non-singular matrices.

We proceed to develop some further consequences of the theorem. First we need the following result:

Theorem 1.6 . A semi-row-reduced square echelon matrix with no zero rows is a diagonal matrix. A row-reduced square echelon matrix with no zero rows is the unit matrix I .

The theorem is an obvious consequence of the definitions.

Theorem 1.7 . A necessary and sufficient condition that the arbitrary $n \times n$ matrix A shall be non-singular is that it shall be row-equivalent to the matrix I_n .

For the proof, we first take as our hypothesis that A is known to be row-equivalent to I_n . In that case, as was noted above, a non-singular matrix E exists, which is a product of elementary matrices, and is such that $EA = I_n$. Premultiplying both sides of this equation by E^{-1} , and then postmultiplying by E , we find that $AE = I_n$. Thus the existence of a matrix E such that $EA = AE = I_n$ has been established, and by definition, this matrix E must be the inverse of A .

Now suppose that A is non-singular and its inverse is a matrix B . By Theorem 1.5, a matrix E exists such that EA is some row-reduced echelon matrix R . Also, $AB = I_n$. Putting these facts together, we find that $EAB = RB$, and $EAB = EI_n$, so $RB = E$. Now if the last row of R were a zero row, then the last row of $RB = E$ would be a zero row also. But this is impossible, because E is non-singular. (No non-singular matrix C can ever have a zero row, because the corresponding row in the product CC_1 for any C_1 would be a zero row, and therefore such a product could never be equal to I .) Therefore since R has no zero rows, by Theorem 1.6, R must be I_n . The proof of the theorem is complete.

Theorem 1.8 . Every non-singular square matrix can be expressed as a product of non-singular elementary matrices. The product can be arranged in the form $A_3 A_1 A_2$ where A_3 is the product of elementary matrices corresponding to elementary row operations of Type (iii); A_1 is a product of elementary matrices

corresponding to elementary row operations of Type (i), and A_2 is a diagonal matrix which is the product of elementary matrices corresponding to elementary row operations of Type (ii).

For the proof, we use Theorems 1.7 and 1.4, together with one of the remarks which followed the proof of Theorem 1.5.

We have (in the notation of that remark) the equation

$$E_2 E_1 E_3 A = I ; \text{ therefore } A = (E_2 E_1 E_3)^{-1} = E_3^{-1} E_1^{-1} E_2^{-1}.$$

This completes the proof.

In the diagonal matrix $A_2 = E_2^{-1}$, the principal diagonal elements are the pivots. That is, the principal diagonal elements are the leading elements of the echelon matrix $E_1 E_3 A$.

We now consider the relationship of Theorem 1.5 to the problem of solving a system of linear equations $A\xi = \eta$, in which A is an arbitrary $n \times m$ matrix (possibly singular), ξ is an unknown $m \times 1$ matrix, and η is a given $n \times 1$ matrix. From Exercise 18 of Section 1.2 we know that a necessary and sufficient condition for ξ to be a solution of $A\xi = \eta$ for a given A and η , is that it shall be a solution of $EA\xi = E\eta$, where E is a product of elementary matrices, ^{premultiplication by} which reduces A to a semi-row-reduced form. It is apparent from this fact that the zero rows in EA will play an important role in the discussion, because if they are not matched by zero rows in $E\eta$, then the system will certainly be unsolvable.

The construction given in the proof of Theorem 1.5 is not ambiguous; it leads to just one semi-row-reduced echelon matrix for a given matrix A , and to just one row-reduced echelon matrix. But it is clearly possible to devise a number

of alternative sequences of elementary row operations which will, in a given case, reduce a given matrix A to semi-row-reduced form. For example, we could start the reduction process with the last non-zero row of A instead of the first. In view of the importance of the number of zero rows in our subsequent discussion, it is of interest to find out whether or not the number of zero rows in a semi-row-reduced matrix R to which a matrix A is row-equivalent, is invariant under different methods of reducing A to such a form. Theorems 1.9 and 1.10 answer this question in the affirmative:

Theorem 1.9 . If R is any semi-row-reduced $n \times m$ matrix, premultiplication of R by an arbitrary non-singular $n \times n$ matrix B can only increase the number of non-zero rows and decrease the number of zero rows.

For the proof, we express B in the form $B = B_3 B_1 B_2$, where the matrices B_3 , B_1 , and B_2 correspond to the matrices A_3 , A_1 and A_2 of Theorem 1.8 . Now first, $B_2 R$ is obtained from R by merely multiplying some of the rows of R by non-zero constants. Therefore this matrix has just the same number of zero rows as R . Also, $B_1 B_2 R$ is only a row-permuted version of $B_2 R$, so it again has the same number of zero rows as R . Of course, $B_1 B_2 R$ is still a semi-row-reduced matrix.

This puts the burden of the proof onto showing what premultiplication by B_3 does to a semi-row-reduced matrix. But elementary row operations of Type (iii) can never reduce the number of non-zero rows in a semi-row-reduced matrix. For let $a^{i_k j_k}$ be a typical leading element in such a matrix. Its

column consists only of zeros except for this element. Therefore essentially the only way to replace such an element with a zero by Type (iii) row operations is to add multiples of the i_k -th row to some other rows and then subtract off a multiple of some one of these rows from the i_k -th. But now some of the zeros previously existing in the j_k -th column have been filled in with non-zero elements, and certainly the number of zero rows has not been increased. The proof of the theorem is complete.

Theorem 1.10 . If A is an arbitrary $n \times m$ matrix and B is any non-singular $n \times n$ matrix such that $BA = R$, where R is a semi-row-reduced matrix, then the number of zero rows in R is independent of the choice of B.

Of course, R will vary with B, but the idea is that the number of zero rows depends only on A.

For the proof, let B_1 and B_2 be non-singular matrices and R_1 and R_2 be semi-row-reduced matrices such that $B_1A = R_1$, $B_2A = R_2$. Then $A = B_1^{-1}R_1$, so $B_2B_1^{-1}R_1 = R_2$. Theorem 1.9 states that R_2 has the same number, or fewer, zero rows than R_1 . On the other hand, $A = B_2^{-1}R_2$, $B_1B_2^{-1}R_2 = R_1$. Theorem 1.9 now states that R_1 has the same number, or fewer, zero rows than R_2 . Therefore R_1 and R_2 have the same number of zero rows.

Thus the number of non-zero rows in any semi-row-reduced matrix to which an $n \times m$ matrix A is row-equivalent, is an

invariant of the method of reduction". We shall call this invariant the rank of the matrix A . It should be noted that the rank can never exceed either n or m . By Theorems 1.6 and 1.7, if the rank of a square $n \times n$ matrix equals n , then the matrix is non-singular.**

If r is the rank of the $n \times m$ matrix A , then $n-r$ is often called the nullity of A .

* As a matter of fact, with somewhat more attention to detail it can be proved that the row-reduced echelon matrix arrived at in Theorem 1.5 (c) is the unique matrix of this kind to which the original matrix A is row-equivalent. Thus it is a canonical form for matrices under row equivalence. We omit the proof, for the result seems to be of only secondary interest for the purposes of this book. See Birkhoff-MacLane (1953), pp 176-177 or Mac Duffee (1947), pp. 35-38. Another canonical form is derived in Exercise 3. at the end of this section. In Chapter 2, a somewhat more stringent uniqueness theorem relative to the triangularization of non-singular matrices will be stated and proved under the name, the "LDU Theorem" (Theorem 2.1).

** We note also here that any echelon matrix can be reduced by elementary row operations to the semi-row-reduced echelon form without disturbing the number of zero rows. Therefore, the number of zero rows in any echelon matrix to which a matrix A is row-equivalent must be an invariant of the method of reduction. In practice this remark often provides the easiest method of determining rank.

There are several other definitions of rank in common use, which for matrices with elements belonging to the number fields under consideration in this book (the real and complex fields) all turn out to be numerically equivalent to the one which we have just given. The theoretical basis for one of the alternative definitions will now be developed. A discussion of the other definitions is postponed to later sections of this chapter. (See Table 1.1 in Section 1.6 for a final summary of equivalent definitions.)

First we must introduce the concept of a minor of a matrix. A minor of a matrix A is any square submatrix obtained from A by suppressing certain rows and certain columns of A . For convenience we include the case in which A is square and no rows and columns are suppressed; the minor then is merely A itself. We shall call the number of rows (or columns) in a minor, the size of the minor.

A principal minor of A is a minor formed by suppressing corresponding rows and columns of A . The principal diagonal of a principal minor consists of elements belonging to the principal diagonal of A . (The concept of a principal minor will not be used in this book until Chapter 2.)

Theorem 1.11 The rank of any matrix equals the number of rows in the non-singular minor of maximal size.

The theorem means that the rank of a matrix A is equal to a number d which has these properties: (1) there is a non-singular minor with d rows in A ; (2) for each $h > d$, every $h \times h$ minor of A is singular. In elementary text-books

the definition of rank is often based on this number d , rather than on the number of non-zero rows in an equivalent row-reduced form.

To prove the theorem, we note first that any elementary row operation on a matrix A is at the same time an elementary row operation (of the same type with essentially the same effect) upon any minor, provided that the rows operated upon are represented in the minor.

Now let M be a minor of A which is known to be non-singular. We slightly alter the construction used in the proof of Theorem 1.5(a) as follows: Perform first a series of Type (iii) row operations on A which reduce M to semi-row-reduced form. There will be no zero rows in this transform of M , which we denote by M_1 . Then add multiples of the rows in A represented in M , to all other rows of A in such a way as to reduce to zero all the elements (except the leading elements of M_1) in all of the columns of A represented in M .

We now proceed to complete the reduction of the entire matrix to semi-row-reduced form by Type (iii) elementary row operations. But no matter how we do this, the rows represented in M will never become zero rows, because nothing that we do will affect the leading elements of M_1 . Therefore, the rank of A is certainly at least equal to the size of M .

But it is easy also to show that there must be at least one non-singular minor of A of size equal to the rank of A . For, suppose that we have transformed A into a row-reduced matrix R by elementary row operations. Consider the minor M_R of R obtained by striking out all of the zero rows in R ,

and all of the columns not containing leading elements. Obviously M_R is itself in row-reduced form, and it is non-singular since it has no zero rows. Its size is equal to the rank of A . But now reverse the transformation which took A into R . The minor M_R goes into a similarly placed minor M_A in A , of the same size as M_R . This minor M_A cannot be singular, because it is row-equivalent to the row-reduced matrix M_R , which has no zero rows.

The proof is now complete.

Exercises.

1. Prove that the inverse of $I - U_{ii} - U_{jj} + U_{ij} + U_{ji}$ $i \neq j$, is $I - U_{ii} - U_{jj} + U_{ji} + U_{ij}$. (Thus this matrix is its own inverse.) Also prove that the inverse of $I + cU_{ij}$, ($c \neq -1$ if $i = j$) is $I - cU_{ij}$.
2. Prove that if a matrix A is non-singular, then the effect of any sequence of elementary row operations can always be accomplished by some sequence of elementary column operations, and vice versa. Show by an example that the proposition is false for singular matrices in general. (Hint: Show that the equation $EA = AX$ has a solution X , and use Theorem 1.8.)
3. Let B be an $n \times n$ non-singular matrix and let A be any $n \times m$ matrix. Prove that the rank of BA is equal to that of A . (Hint: An equivalent statement would be that elementary row operations cannot affect the rank of a matrix. Let E_B be a product of elementary matrices such that $E_B B = I_n$. Let E_A be a product of elementary matrices such that $E_A A$ is in row reduced form. Then $E_A(E_B B)A = E_A I_n A = E_A A$. Thus $E_A E_B$ is a product of elementary matrices which reduce BA to

arbitrary $n \times m$ matrix, and let C be a non-singular $m \times m$ matrix. Prove that the rank of BAC is equal to that of A .

8. Show that the reduced form introduced in Exercise 4. is unique, in the sense that if by some sequence of elementary row and column operations, the matrix A of rank r is reduced to the form J_s , then necessarily $s = r$. (Hint: Use Exercise 7.)

9. Prove that a matrix A and its transpose A' have the same rank. (Hint: If $E_1 A E_2 = J_r$, where J_r is the special matrix introduced in Exercise 4., then $E_2' A' E_1' = J_r'$. The proof is completed by inspection of J_r' and reference to Exercise 6. Another simple proof can be given on the basis of Exercise 15 of Section 1.2, and Theorem 1.11.)

10. Let $\text{rk}(A)$ denote the rank of A . Prove that $\text{rk}(AB) \leq \text{rk}(A)$, and simultaneously $\text{rk}(AB) \leq \text{rk}(B)$, where of course A and B are assumed to be conformable. (Hint: Let E_1 and E_2 be products of elementary matrices such that $E_1 A$ and $E_2 B'$ are in semi-row-reduced form. By Exercise 7, the rank of $E_1 A B E_2'$ is the same as that of AB . Now count up the number of rows and columns in $E_1 A B E_2'$ which certainly must be zero, and consider what would happen if this matrix were reduced to the form appearing in Exercise 4.)

11. Prove that if A is any $n \times m$ matrix, and if it is known that a matrix B exists such that $AB = I$, then B is the inverse of A . Similarly prove that if a matrix B_1 exists such that $B_1 A = I$, then B_1 is the inverse of A . (Hint: By Exercise 10., the ranks of both A and B cannot be less than n .)

12. Show by an example that the rank of AB is not necessarily equal to that of BA .

1.4. Systems of linear equations. We are now ready to proceed with the problem of solving $A\xi = \eta$. We take up the case in which $\eta = 0$ first*. The system $A\xi = 0$ is called homogeneous.

Theorem 1.12. Let A be an arbitrary $n \times m$ matrix and let r be its rank. A necessary and sufficient condition for there to be non-zero solutions $\xi = (x^1, x^2, \dots, x^m)$ to the system $A\xi = 0$ is that $r < m$. The totality of solutions if $r < m$ is obtained by letting a certain $m - r$ elements of ξ run through all the numbers of the field to which the elements of A belong; for each assignment of the $m - r$ elements of ξ , the remaining r elements are uniquely determined by the equations of the system $A\xi = 0$.

The proof is given by transforming A to the row-reduced echelon form by premultiplying by a product E of elementary matrices. The system then goes over into the equivalent one $R\xi = 0$, where $R = EA$ is a row-reduced echelon matrix. At this point, the reader may find it helpful to look at a sample 4×7 row-reduced echelon matrix:

$$\begin{bmatrix} 0 & 1 & a^{13} & 0 & 0 & a^{16} & a^{17} \\ 0 & 0 & 0 & 1 & 0 & a^{26} & a^{27} \\ 0 & 0 & 0 & 0 & 1 & a^{36} & a^{37} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

* The symbol 0 here stands for an $n \times 1$ zero matrix.

The corresponding system of homogeneous linear equations is

$$\begin{aligned} 0x^1 + 1x^2 + a^{15}x^3 + 0x^4 + 0x^5 + a^{16}x^6 + a^{17}x^7 &= 0 \\ 0x^1 + 0x^2 + 0x^3 + 1x^4 + 0x^5 + a^{26}x^6 + a^{27}x^7 &= 0 \\ 0x^1 + 0x^2 + 0x^3 + 0x^4 + 1x^5 + a^{36}x^6 + a^{37}x^7 &= 0 \\ 0x^1 + 0x^2 + 0x^3 + 0x^4 + 0x^5 + 0x^6 + 0x^7 &= 0. \end{aligned}$$

From this special case the general situation can easily be visualized. If $r < m$, every element of $\xi = (x^1, \dots, x^m)$ which is not multiplied in some row by a leading element of R , can be assigned arbitrarily. Thereafter it is trivial to solve for the remaining elements of ξ , because each such element appears with a non-zero coefficient in one and only one equation.

But if $r = m$, then the echelon character of R implies that R may be partitioned as follows:

$$R = \begin{bmatrix} I_m \\ 0 \end{bmatrix}$$

The first equation of the system $R\xi = 0$ is $x^1 = 0$, the second is $x^2 = 0$, and so forth. Thus there can be no non-zero solutions in this case.

Corollary. If A is an $n \times m$ matrix, with $n < m$, then the system $A\xi = 0$ always has non-zero solutions.

For then the rank of A is surely less than m .

Theorem 1.13 . Let A be an arbitrary $n \times m$ matrix. Two alternative necessary and sufficient conditions for the existence of solutions to the system $A\xi = \eta$ are that

(a) The rank of the augmented matrix $[A; \eta]$ is the same

as that of A ; or

(b) The equation $n'\theta = y^1t^1 + y^2t^2 + \dots + y^nt^n = \theta'n = 0$ holds true for every solution $\theta = (t^1, t^2, \dots, t^n)$ of the transposed homogeneous system $A'\theta = 0$.

For part (a), let E be a product of elementary matrices which reduces the $n \times (m+1)$ matrix $\begin{bmatrix} A' & \eta \end{bmatrix}$ to row-reduced echelon form. The resulting matrix is $\begin{bmatrix} EA & E\eta \end{bmatrix}$. We know that each solution of $A\xi = \eta$ is also a solution of $EA\xi = E\eta$, and vice versa. Moreover, EA is itself in row-reduced echelon form, and so the number of zero rows in EA is equal to the rank of A.

There are now two possibilities. The first is that $E\eta$ has a non-zero element in ^{some} $\wedge$ row in which all the elements of EA are zero. In this case there obviously can be no solutions of $EA = E\eta$. Also, in this case, the rank of $\begin{bmatrix} A' & \eta \end{bmatrix}$ is greater than the rank of A.

The other possibility is that $E\eta$ has no non-zero element below the last non-zero row of EA. The rank of $\begin{bmatrix} A' & \eta \end{bmatrix}$ is then equal to that of A. In this case, solutions can be found exactly as described in the proof of Theorem 1.12. The proof of part (a) is complete.

We remark that the proof shows that the totality of solutions of $A\xi = \eta$ is obtained just as in the homogeneous case, by letting a certain $m - r$ elements of the typical solution ξ run through all the numbers of the field to which the elements of A and η belong, where r is the rank of A.

We now turn to part (b). In the first place, it is easy to prove that if solutions of $A\xi = \eta$ do exist, and if θ is any solution of $A'\theta = 0$, then $\theta'\eta = 0$. For taking transposes, the system $A'\theta = 0$ is seen to be the same as $\theta'A = 0$, where now the zero on the right side represents the $1 \times m$ zero matrix. Let ξ be a solution of $A\xi = \eta$. Premultiplying by θ' , we obtain $\theta'A\xi = \theta'\eta$. But $\theta'A = 0$, so $\theta'A\xi = 0 = \theta'\eta$.

The converse is somewhat more subtle. Let E be a product of elementary matrices such that $EA = R$, where R is a row-reduced $n \times m$ echelon matrix. If γ' is a $1 \times n$ row matrix which satisfies the system $\gamma'R = 0$ (which is the same as $(\gamma'E)A = 0$), then $\theta' = \gamma'E$ is a solution of $\theta'A = 0$. That is, the totality of solutions θ' of $\theta'A = 0$ certainly includes all row matrices of the form $\theta' = \gamma'E$, where γ' runs through the totality of solutions $\gamma'R = 0$.

Our hypothesis now is that the $n \times 1$ matrix η satisfies the equation $\theta'\eta = 0$ for every solution θ' of the system $\theta'A = 0$. Therefore in particular, $(\gamma'E)\eta = 0$ for every solution γ' of $\gamma'R = 0$.

We shall now show that the totality of solutions $\gamma' = [g^1 \ g^2 \ \dots \ g^n]$ of $\gamma'R = 0$ is represented by

$$(1.4.1) \quad \gamma' = \left[\overbrace{0 \ 0 \ \dots \ 0}^r \ g^{r+1} \ g^{r+2} \ \dots \ g^n \right],$$

where the last $n - r$ elements run through all the numbers of the number field under consideration. (If the reader considers this to be self-evident, he can skip the next paragraph.)

The various equations of the system $\gamma'R = 0$ correspond respectively to the columns of R . The situation can be

visualized more readily by means of an example. Let R be

$$\begin{bmatrix} 1 & 0 & a^{13} & 0 & a^{15} \\ 0 & 1 & a^{23} & 0 & a^{25} \\ 0 & 0 & 0 & 1 & a^{35} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The corresponding system $Y'R = 0$ (written out vertically to facilitate visualization of the general situation) is

$$\begin{array}{ccccc} 1g^1 & 0g^1 & a^{13}g^1 & 0g^1 & a^{15}g^1 \\ \div 0g^2 & \div 1g^2 & \div a^{23}g^2 & \div 0g^2 & \div a^{25}g^2 \\ \div 0g^3 & \div 0g^3 & \div 0g^3 & \div 1g^3 & \div a^{35}g^3 \\ \div 0g^4 & \div 0g^4 & \div 0g^4 & \div 0g^4 & \div 0g^4 \\ \hline = 0 & = 0 & = 0 & = 0 & = 0 \end{array}$$

From this display it can be inferred that in the general situation, since R is a row-reduced echelon matrix, there will always be an equation in the system which implies that $g^1 = 0$, then another which implies that $g^2 = 0$, then another which implies that $g^3 = 0$, and so on, until the first r components of Y' have been determined. Thus any solution Y' of $Y'R = 0$ is of the form (1.4.1), in which g^{r+1} , g^{r+2} , ..., g^n can be assigned entirely arbitrarily, because they appear only with zero coefficients in the system $Y'R = 0$.

Now we return to the equation $(Y'E)\eta = 0 = Y'(E\eta)$ which (by hypothesis) is satisfied by the solutions of $Y'R = 0$. The equation certainly imposes no restrictions on the first r elements of the column matrix $E\eta$, because they are matched with zero elements in Y' . But the last $n - r$ elements of $E\eta$

must be such that the sum of their respective products into arbitrary numbers $g^{r+1}, g^{r+2}, \dots, g^n$ must be zero. This can happen only if all of the last $n - r$ elements of $E\eta$ are zeros, because we could in particular choose $g^{r+1} = 1, g^{r+2} = 0, \dots, g^n = 0$, then $g^{r+1} = 0, g^{r+2} = 1, g^{r+3} = 0, \dots, g^n = 0$, and so forth.

But if the last $n - r$ elements of $E\eta$ are zeros, then of course the rank of $[A | \eta]$ is equal to r , because premultiplying $[A | \eta]$ by E reduces it to a row-reduced echelon form having the same number of non-zero rows as $EA = R$. Thus by part (a), there must be solutions to the system $A\xi = \eta$. The proof of the theorem is complete.

The reader may notice at this point that we have now established all the assertions in Theorem 1.1, the fundamental theorem, except for the statements relating to the linear independence of the solutions of $A\xi = 0$ and $A'\theta = 0$ in the singular case. These will be taken care of in the next section.

Exercises

1. Prove that if the rank of the $n \times m$ matrix A is r , then the totality of solutions of the system $A'\theta = 0$ is obtained by letting a certain ^{set of} $n - r$ elements of θ run through the field of numbers to which the elements of A belong. (Hint: Refer to Theorem 1.12 and Exercise 9 of Section 1.3.)
2. Prove that all the solutions of the system $A\xi = \eta$ are given by the formula $\xi = \xi_0 + \zeta$, where ξ_0 is any particular solution of $A\xi = \eta$ and ζ ranges through all the solutions of $A\xi = 0$. (Hint: Obviously each column matrix $\xi_0 + \zeta$ is indeed a so-

lution of $A\xi = \eta$, by direct substitution. Now use the remark which follows the proof of Theorem 1.13 (a), together with the fact that if an element x of a variable column matrix ξ is supposed to range through a field of numbers, this can be accomplished by letting $x = x_0 + z$, where x_0 is any constant in the field and z runs through the field.)

3. Prove that if the rank of the $n \times m$ matrix A is n , then the system $A\xi = \eta$ always has solutions for any η .

4. Prove that if A is $n \times n$ and if for some one η the system $A\xi = \eta$ is known to have a unique solution, then A is necessarily non-singular.

5. Given any $n \times 1$ column matrix ξ , show that an $n \times n$ non-zero matrix A can always be found such that $A\xi = 0$.

6. Prove that if A is $n \times m$ and the rank of A is m , then if a solution of the system $A\xi = \eta$ exists at all, it is unique. (Note: Only the case $n > m$ has any novelty at this point, since the case $n = m$ is merely the case of a non-singular A .)

7. Let A and B be $n \times n$ matrices and let η be an $n \times 1$ column matrix. Prove that if a column matrix ξ exists such that simultaneously $A\xi = \eta$ and $B\xi = \eta$, then $A - B$ must be singular.

8.* An $n \times n$ matrix $A = [a^{ij}]$ is said to have a dominant diagonal if $|a^{ii}| > \sum_{j \neq i} |a^{ij}|$, $i = 1, \dots, n$. Prove

that if a matrix A has a dominant diagonal, then A is neces-

* For references concerning the result stated in this exercise, see Taussky (1949).

sarily non-singular. (Hint: Assume the contrary. Then a non-zero solution $\xi = (x^1, \dots, x^n)$ exists for the system $A\xi = 0$. Let x^{i_1} be an element of the solution which has the largest absolute value. Then $a^{i_1 i_1} x^{i_1} = -a^{i_1 1} x^1 - \dots - a^{i_1, i_1-1} x^{i_1-1} - a^{i_1, i_1+1} x^{i_1+1} - \dots - a^{i_1, n} x^n$. Now use the fact that the absolute value of a sum of complex numbers cannot exceed the sum of the absolute values of the numbers, to reach a contradiction with the dominant diagonal property.)

9.* Another way to characterize the dominant diagonal property is to write it as a set of equations as follows:

$\sum_{j \neq i} |a^{ij}| = r_i |a^{ii}|$, $0 \leq r_i < 1$, $i = 1, \dots, n$. Let A be a complex matrix which has this property, and let $A^{-1} = B = [b^{ij}]$. Prove that B has the property that

$$(1.4.2) \quad |b^{ij}| \leq r_i |b^{jj}|, \quad i \neq j,$$

$$(1.4.3) \quad b^{ii} = \frac{1}{a^{ii} + p_i \hat{r}_i r_i |a^{ii}|}, \quad |p_i| \leq 1,$$

$$i, j, = 1, 2, \dots, n, \text{ where } \hat{r}_i = \max_{j \neq i} r_j.$$

$$(\text{Note: Equation (1.4.3) implies that } |b^{ii}| \leq \left| \frac{1}{a^{ii}} \right| \left| \frac{1}{1 - \hat{r}_i r_i} \right|.)$$

(Hint: The proof is based on the fact that by the definition of inverse, $\sum_{k=1}^n a^{ik} b^{kj} = 0$ if $j \neq i$, and this summation equals one if $j = i$. Let b_j be the maximum of $|b^{kj}|$ for $k = 1, \dots, n$. The strategy for (1.4.2) is to prove that $b_j = |b^{jj}|$, at the same time showing that $|b^{ij}| \leq r_i b_j$. We have $|a^{ii}| |b^{ij}| = \left| - \sum_{k \neq i} a^{ik} b^{kj} \right| \leq b_j \sum_{k \neq i} |a^{ik}| \leq b_j r_i |a^{ii}|$.

We leave the remainder of the proof of (1.4.2) to the reader.

As for (1.4.3), we have

$$\left| a^{ii} b^{ii} - 1 \right| \leq \sum_{k \neq i} \left| a^{ik} b^{ki} \right| \leq \sum_{k \neq i} \left| a^{ik} r_k b^{ii} \right| \leq \left| b^{ii} \right| r_i \sum_{k \neq i} \left| a^{ik} \right|.$$

The reader should have no difficulty in proceeding from here.)

10.* Let $A = [a^{ij}]$ be a complex $n \times n$ matrix and let η be a complex $n \times 1$ matrix with elements $y^1, y^2, \dots, y^n$. Further, let A and η have the following property:

$$\left| a^{ii} \right| > \left| \frac{y^i}{r} \right| + \sum_{j \neq i} \left| a^{ij} \right|, \quad i = 1, \dots, n, \text{ where}$$

r is some positive number. Prove that the system of linear equations represented by $A\xi = \eta$ has a unique solution

$\xi = (x^1, x^2, \dots, x^n)$ such that $|x^i| < r$, $i = 1, \dots, n$.

(Hint: Of course the existence and uniqueness of the solution is assured by Exercise 8. Let $|x^k|$ be the maximum of $|x^1|, |x^2|, \dots, |x^n|$, and suppose that $|x^k| \leq r$, in contradiction to the conclusion stated above. Consider the equation $-a^{kk}x^k = -y^k + \sum_{j \neq k} a^{kj}x^j$.

Then $|a^{kk}||x^k| \leq |y^k| + |x^k| \sum_{j \neq k} |a^{kj}|$. Divide both sides by

$|x^k|$ and use the fact that $|x^k| \geq r$ to obtain a contradiction

with the hypothesis.)

11.** Let $A = [a^{ij}]$ be a complex $n \times n$ matrix such that

$|a^{ii}| \leq \sum_{j \neq i} |a^{ij}|$, $i = 1, \dots, n$, with equality in at most

$n - 1$ cases. Assume further that there is no matrix P con-

* Price (1951)

** Taussky (1949)

sisting of a product of elementary matrices corresponding to Type (i) elementary row or column operations such that

$$PAP = \left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline 0 & A_{22} \end{array} \right]$$

where A_{11} and A_{22} are square matrices. Show that these hypotheses imply that A is non-singular. (Hint: As usual, assume the contrary and let $|x^{i_1}|$ be an element of maximal absolute value of a solution of $AX = 0$. By considering a particular equation of the system which corresponds to a row of A for which $|a^{ii}| > \sum_{j \neq i} |a^{ij}|$, it can be shown that $|x^{i_1}| > |x^i|$ for at least one i . Now consider each of the equations corresponding to the row indices $i_1, i_2, \dots, i_s$ such that $|x^{i_1}| = |x^{i_2}| = \dots = |x^{i_s}|$. Either of two things can happen: (a) From any one of these equations a contradiction of the diagonal condition of the hypothesis can be inferred, or (b) the matrix must be of the excluded type.)

12.* Let $A = [a^{ij}]$ be a real $n \times n$ matrix such that $a^{ii} \geq 0$, $a^{ij} \leq 0$, $i, j = 1, \dots, n$, $i \neq j$. Further let A satisfy the hypotheses of Exercise 11. Then A is singular if and only if $\sum_{j=1}^n a^{ij} = 0$, $i = 1, \dots, n$. (Hint: If the sum of the elements in each row is zero, then by adding the first $n - 1$ columns to the last one, which can be done by elementary column transformations of Type (iii), we obtain an equivalent matrix with a column of zeros. So this matrix must be singular. Conversely, if A is singular, proceed as

* Markov (1908). The Hint suggests the proof given by Taussky (1949).

in the proof of Exercise 11. to show that all the elements of any solution of $A\xi = 0$ are equal in absolute value.)

^c1.5 Vectors. The vectors employed in physics to represent forces and other physical concepts can be represented by ordered triples of real numbers. Generalizing this idea, we shall call any real or complex matrix with just one column or just one row a vector.

The elements of a vector are often called its com-ponents. If the number of components of a given vector is, say, n , and if it is desirable to emphasize this fact, we shall refer to the vector as an n -vector.

As was stated above, a vector may be either a column matrix or a row matrix. However, in this book it will be convenient to adopt the convention that each vector is to be regarded as a column matrix unless there is a specific indication to the contrary. The indication that a vector should be thought of as a row matrix will take the form of a "prime" (indicating transposition) applied to the symbol for the vector. For example, if a vector α with components $a^1, a^2, \dots, a^n$ is introduced in a discussion without further qualification, it is to be understood that we are referring to the column matrix

$$\alpha = (a^1, a^2, \dots, a^n) = \begin{bmatrix} a^1 \\ a^2 \\ \cdot \\ \cdot \\ \cdot \\ a^n \end{bmatrix} .$$

The notation α' then will be used to refer to the row matrix

$$\alpha' = [a^1 \quad a^2 \quad \dots \quad a^n].$$

The algebra of vectors is formally identical with the algebra of matrices, specialized to the case of column or row matrices. A special case of multiplication which is of some importance is that of a product of the type $\alpha' \bar{\beta}$, where α and β both have the same number of components. This is called the scalar product, or inner product, or dot product, of the two vectors. It is often denoted by the symbol (α, β) . In terms of the respective components a^i and b^i , $i = 1, \dots, n$, of α and β , the formula is as follows:

$$\alpha' \bar{\beta} = (\alpha, \beta) = a^1 \bar{b}^1 + a^2 \bar{b}^2 + \dots + a^n \bar{b}^n = (\bar{\beta}, \bar{\alpha}) = \bar{\beta}' \alpha$$

Note that if α and β are real, $(\alpha, \beta) = (\beta, \alpha)$. The reader should also notice that although $\alpha' \bar{\beta}$ is a scalar, $\alpha \bar{\beta}'$ is a square hermitian $n \times n$ matrix. (See Exercise 2 at the end of this Section.)

The square root of (α, α) is sometimes called the length of α when α is real. The notation $|\alpha|$ is sometimes used for the length of α .

Some properties of scalar products will be developed in Exercise 1 at the end of this section.*

The most important concept relating to vectors is that of linear independence. The set of n -vectors $\alpha_1, \alpha_2,$

* We shall not use the special notation (α, β) for scalar products outside of this section. For the purposes of this book it seems to have no advantage over the notation $\alpha' \bar{\beta}$.

..., α_m are said to be linearly independent if the system of n simultaneous linear equations represented by the vector equation $c_1\alpha_1 + c_2\alpha_2 + \dots + c_m\alpha_m = 0$ has the unique solution $c_1 = c_2 = \dots = c_m = 0$.

Vectors which are not linearly independent are naturally said to be linearly dependent. If $\alpha_1, \dots, \alpha_m$ are linearly dependent, then there exists scalars $c_1, c_2, \dots, c_m$ not all zero such that

$$c_1\alpha_1 + c_2\alpha_2 + \dots + c_m\alpha_m = 0.$$

(Of course it is understood that scalars c_j and the components of the vectors lie in the same field.)

A set of n linearly independent n -vectors which will appear from time to time on these pages, is the set of so-called unit vectors

$$\begin{aligned} \epsilon_1 &= (1, 0, 0, \dots, 0) \\ \epsilon_2 &= (0, 1, 0, \dots, 0) \\ \epsilon_3 &= (0, 0, 1, \dots, 0) \\ &\vdots \\ \epsilon_n &= (0, 0, 0, \dots, 1) \end{aligned}$$

We shall always use epsilons with subscripts to denote these vectors. The fact that they are linearly independent is obvious from the definition of linear independence.

Some fairly immediate general consequences of the definition of linear independence are contained in the following theorems.

Theorem 1.14 . Any subset of a set of linearly independent vectors is itself a linearly independent set of vectors.

Theorem 1.15 . A set of linearly independent vectors can never contain a zero vector.

Theorem 1.16 . If a set of vectors $\alpha_1, \alpha_2, \dots, \alpha_m$ is linearly dependent, then any larger set including $\alpha_1, \alpha_2, \dots, \alpha_m$ as a subset is also a linearly dependent set.

Theorem 1.17 . If the set of n-vectors $\alpha_1, \alpha_2, \dots, \alpha_m$ is a linearly independent set, and if B is an arbitrary non-singular $n \times n$ matrix, then $B\alpha_1, B\alpha_2, \dots, B\alpha_m$ are also linearly independent. If $\alpha_1, \alpha_2, \dots, \alpha_m$ are linearly dependent, then $C\alpha_1, C\alpha_2, \dots, C\alpha_m$ are also linearly dependent, where C is any $n \times n$ matrix.

The proof of the second statement in Theorem 1.17 is given by multiplying each side of the equation $C_1\alpha_1 + C_2\alpha_2 + \dots + C_m\alpha_m = 0$ by C . The first statement is thereafter proved by assuming that $B\alpha_1, B\alpha_2, \dots, B\alpha_m$ are linearly dependent, and using the second statement with $C = B^{-1}$ to reach a contradiction with the hypothesis.

We now develop a connection between linear independence and the rank of a matrix.

Clearly the columns of an $n \times m$ matrix can be thought of as a set of n -vectors. A similar remark applies to the rows, with m substituted for n . For every matrix, there must be a unique integer k such that k particular columns are linearly independent, but at the same time either $k = n$ or

every set of $k + 1$ columns always forms a linearly dependent set of vectors. We naturally say that k is "the" number of linearly independent columns of the matrix. The number of linearly independent rows is similarly defined.

Theorem 1.18 . Let A be an arbitrary rectangular matrix.
The rank of A is identical with the number of linearly inde-
pendent columns. It is also identical with the number of
linearly independent rows.

The theorem is an immediate consequence of the theory of homogeneous simultaneous linear equations as developed in Section 1.4 . We first consider the columns of A , and recall that a set of columns $\alpha_j = (a^{1j}, a^{2j}, \dots, a^{nj})$, $j = j_1, j_2, \dots, j_k$, are linearly independent if and only if the system of equations

$$\begin{aligned}
 & a^{1j_1} c_{j_1} + a^{1j_2} c_{j_2} + \dots + a^{1j_k} c_{j_k} = 0 \\
 & a^{2j_1} c_{j_1} + a^{2j_2} c_{j_2} + \dots + a^{2j_k} c_{j_k} = 0 \\
 & \cdot \\
 & \cdot \\
 & \cdot \\
 & a^{nj_1} c_{j_1} + a^{nj_2} c_{j_2} + \dots + a^{nj_k} c_{j_k} = 0
 \end{aligned}
 \tag{1.5.1}$$

has the

unique solution $c_{j_1} = c_{j_2} = \dots = c_{j_k} = 0$.

If the rank of A is r , then by Theorem 1.11, there exists a certain non-singular minor of A of size r . Let the columns of A represented in this minor be $\alpha_{j_1}', \alpha_{j_2}', \dots, \alpha_{j_r}'$. Set up the system (1.5.1) for these vectors with $k = r$. The rank

of the matrix of the system (1.5.1) must be r , by Theorem 1.11. Therefore by Theorem 1.12, the only solution is $c_{j_1} = c_{j_2} = \dots = c_{j_r} = 0$, so the vectors $\alpha_{j_1}, \dots, \alpha_{j_r}$ must be linearly independent.

The same line of reasoning shows that every set of $r + 1$ columns must be linearly dependent. For if the system (1.5.1) is set up for any set of $r + 1$ columns, then the rank of its matrix must be less than $r + 1$, because otherwise, by Theorem 1.11, it would have a non-singular minor of size $r + 1$, and then also would A have such a non-singular minor. But if the rank is less than $r + 1$, then by Theorem 1.12 the system has non-zero solutions.

To discuss rows instead of columns, we have now merely to observe that the rows of A are the columns of A' , and by Exercise 9 of Section 1.3, the ranks of A and A' are equal. The proof is complete.

The number of linearly independent rows in a matrix is sometimes called the row rank. The number of linearly independent columns is similarly called the column rank.*

We now discuss the implications of Theorem 1.18 as regards the number of linearly independent solutions of a system of linear equations.

We first need one simple corollary of the theorem.

Corollary . A set of n -vectors $\alpha_1, \alpha_2, \dots, \alpha_m$, may contain n linearly independent vectors, but never more than n .

* See e. g. MacDuffee (1943), p. 33, p. 42 .

For the rank of the $n \times m$ matrix $A = [\alpha_1 \mid \alpha_2 \mid \dots \mid \alpha_m]$ may be equal to n , but A has only n rows, so its rank can be no greater than n .

Let us now consider the case of the solutions of a system of linear equations, say $A\xi = \eta$, in which A is $n \times m$ and of rank r . We shall study the homogenous case first, in which $\eta = 0$. A glance at the proof of Theorem 1.12 (and particularly at the examples displayed there) reveals that in a typical solution ξ , $m - r$ of the components of ξ are assignable arbitrarily, and then each of the remaining components is determined as a linear function of the arbitrarily assigned ones.* To be more specific, let us relabel the components of a solution ξ as follows: Let $x_0^1, x_0^2, \dots, x_0^{n-r}$ denote the arbitrarily assignable components, and let $x_1^1, x_1^2, \dots, x_1^r$ denote the components to be determined from them. Then the relationship between the first set and second set is of the form

$$x_1^i = b^{i1} x_0^1 + b^{i2} x_0^2 + \dots + b^{i, m-r} x_0^{m-r}, \quad i = 1, 2, \dots, r.$$

*

To simplify this exposition, we are tacitly assuming that the components of ξ which are multiplied by leading elements of the row-reduced echelon matrix used in the proof of Theorem 1.12 are treated as dependent variables, and the $m - r$ other components are treated as independent variables. In specific cases it is not always necessary to set up such an order relationship, but this will not affect our conclusions.

The coefficients b^{ij} of course do not depend on the components x_0^i .

Suppose now that we have determined k presumably different solutions of the system $A\xi = 0$. We designate these by second subscripts on the symbols x_0^i and x_1^i . The matrix having these solutions as columns has the following appearance, possibly after a permutation of rows:

$$X = \begin{bmatrix} x_{01}^1 & \cdots & x_{0k}^1 \\ x_{01}^2 & \cdots & x_{0k}^2 \\ \vdots & & \vdots \\ x_{01}^{m-r} & \cdots & x_{0k}^{m-r} \\ b^{11}x_{01}^1 + \cdots + b^{1,m-r}x_{01}^{m-r} & \cdots & b^{11}x_{0k}^1 + \cdots + b^{1,m-r}x_{0k}^{m-r} \\ \vdots & & \vdots \\ b^{r1}x_{01}^r + \cdots + b^{r,m-r}x_{01}^{m-r} & \cdots & b^{r1}x_{0k}^r + \cdots + b^{r,m-r}x_{0k}^{m-r} \end{bmatrix}$$

From the above display we see at once that X is row-equivalent under elementary row transformations of Type (iii) to a matrix whose first $m-r$ rows are the same as those of X , but whose last r rows are zero rows. So the rank of X is merely the

number of linearly independent $(m-r)$ -vectors $(x_{0j}^1, x_{0j}^2, \dots, x_{0j}^{m-r})$ in the set of solutions. We know from the Corollary of Theorem 1.18 that this number cannot be higher than $m-r$, although it can indeed be that high if the assignable components are suitably chosen. It will be remembered that they can be chosen quite arbitrarily.

In the non-homogeneous case, the questions of solvability and number of solutions are resolved, as in the homogeneous case, by using elementary row transformations to throw the system $A\xi = \eta$ into the form $R\xi = \beta$, where R is a semi-row-reduced echelon matrix. For solutions to exist at all, the vector β must have zero components corresponding to the zero rows of R . By imagining the zeros on the right-hand side of the sample system of homogeneous equations displayed after Theorem 1.12 to be replaced by the components of β , the reader can easily visualize the general situation. In a typical solution, $m-r$ components, say $x_0^1, x_0^2, \dots, x_0^{m-r}$ can be assigned arbitrarily and then the others, say $x_1^1, x_1^2, \dots, x_1^r$ are determined from them by relations of the form

$$x_1^i = b^i + b^{i1} x_0^1 + b^{i2} x_0^2 + \dots + b^{i, m-r} x_0^{m-r}, \quad i = 1, 2, \dots, r.$$

Thus the appropriate modification to the matrix X introduced above in the homogeneous case, consists in adding the constant b^i to each element in the i -th row, $i = m-r+1, \dots, m$. We denote the matrix obtained from X in this way by Y . Now how high can the rank of Y get?

To find the answer, we eliminate (i.e., reduce to zero) by elementary row operations all the terms involving x_0 's in

the rows of Y after the $(m-r)$ -th, just as before. The last r elements in each column of Y are now $b^1, b^2, \dots, b^r$. Then by elementary row operations of Type (iii), all the rows after the $(m-r)$ -th, except possibly one of them, are transformed into zero rows. We end up with a matrix with at most $m-r + 1$ non-zero rows, so its rank and the number of linearly independent solutions cannot exceed $m-r + 1$.

Its rank can, however, equal $m-r + 1$ if the assignable components are chosen properly, as we shall now show. The numbers b^i are the components of the transform of a presumably non-zero vector η by elementary row operations. (If η were zero in the system $A\xi = \eta$, then we would have the homogeneous case.) Thus the numbers b^i cannot all be zero. Therefore one of the last r rows of the transformed matrix Y is a non-zero row. If we allow Y to have $m-r + 1$ columns, and take the vectors $(x_{0j}^1, \dots, x_{0j}^{m-r})$, $j = 1, \dots, m-r$, to be the set of unit vectors, and take $(x_{0, m-r+1}^1, \dots, x_{0, m-r+1}^{m-r})$ to be a zero vector, then Y is row-equivalent to the partitioned matrix

$$Y_1 = \left[\begin{array}{c|c} I_{m-r} & 0 \\ \hline 0 & b \\ \hline 0 & 0 \end{array} \right],$$

where b stands for some one of the non-zero components of β .

(In arriving at this form, we reduced as many elements as possible to zero in the lower part of the matrix by subtracting off multiples of the rows in the top part from the one non-zero row in the lower part.)

The rank of $\bar{Y}_1$ is clearly equal to $m-r+1$, so we have shown that it is possible for the number of linearly independent solutions to equal $m-r+1$.

The results of the several preceding paragraphs can be summarized as follows:

Theorem 1.19 . Let A be an $n \times m$ matrix of rank r . The homogeneous system of equations $A\xi = 0$ has $m-r$ linearly independent solutions. They can be obtained by using in turn for the $m-r$ assignable components of each solution, the respective components of any set of $m-r$ linearly independent $(m-r)$ -vectors. In particular, the set of unit $(m-r)$ -vectors $e_1, e_2, \dots, e_{m-r}$ can be used.

The non-homogeneous system $A\xi = \eta, \eta \neq 0$, has $m-r+1$ linearly independent solutions. They can be obtained by using in turn for the $m-r$ assignable components, the respective components of any set of $m-r+1$ $(m-r)$ -vectors which consists of $m-r$ linearly independent $(m-r)$ -vectors together with a zero $(m-r)$ -vector.

At the end of the last Section, it was observed that all the assertions in Theorem 1.1, the fundamental theorem, had been established at that point except for the statements regarding linear independence of solutions in the singular case. This gap has now been filled, and the proof of Theorem 1.1 is now complete.

The reader has no doubt noticed that the proof was "constructive", in the sense that it consisted of furnishing a recipe for finding the solutions in a finite number of steps.

The proof we have given of Theorem 1.1 is in fact the basis for two of the most successful computational programs for solving linear systems numerically. These are presented under METHOD I in the next chapter.

Exercises.

1. Establish the following properties of scalar products:

(a) $(\alpha, \alpha) \geq 0$ if $\alpha \neq 0$.

(b) $(\alpha_1 + \alpha_2, \beta) = (\alpha_1, \beta) + (\alpha_2, \beta)$.

(c) If α and β are real vectors, then $(\alpha, \beta)^2 \leq (\alpha, \alpha)(\beta, \beta)$, with equality if and only if α and β are linearly dependent. (Note: Property (c) is known as the Schwarz inequality, or Cauchy-Schwarz inequality. Hint: To prove it, notice that

$$\sum_{i=1}^n (a^i)^2 \sum_{i=1}^n (b^i)^2 - \left(\sum_{i=1}^n a^i b^i \right)^2 = \sum_{i=1}^n \sum_{j < i} (a^i b^j - a^j b^i)^2 \quad .)$$

2. If α and β are two complex n -vectors, show that $\alpha \beta'$ is a hermitian matrix.

3. Let ξ be a real n -vector and let A and B be any two real $n \times n$ matrices. Prove that $\xi'(A'B + B'A)\xi = 2(A\xi, B\xi)$.

4. Prove that a square matrix is non-singular if and only if all of its columns (or rows) are linearly independent.

5. Let A be an arbitrary $n \times m$ matrix and let B be a non-singular $n \times n$ matrix. Prove that the number of linearly independent rows or columns of $C = BA$ is the same as the number of linearly independent rows or columns of A . (Hint: Refer to Exercise 3 of Section 1.3.)

6. Show that if A has m columns and B has m rows, and if $AB = 0$, then $\text{rk}(A) + \text{rk}(B) \leq m$. (Hint: The columns of B are

each solutions of the system $A\xi = 0$. What therefore is the largest number of linearly independent columns that B can have?)

7. Prove Sylvester's Law of Nullity: If A has m columns and B has m rows, then $\text{rk}(A) + \text{rk}(B) - m \leq \text{rk}(AB) \leq \min \{ \text{rk}(A), \text{rk}(B) \}$.

(Hint: The right-hand inequality is only a restatement of Exercise 10 of Section 1.3. For the left side, notice first that $B\xi = 0$ has a set of $m - \text{rk}(B)$ linearly independent solutions which are all solutions of $AB\xi = 0$; and $AB\xi = 0$ has $m - \text{rk}(AB)$ linearly independent solutions. By examining the statement of Theorem 1.19, it will be seen that a particular set of $m - \text{rk}(AB)$ linearly independent solutions of $AB\xi = 0$ can be chosen so as to include a set of $m - \text{rk}(B)$ linearly independent solutions of $B\xi = 0$. This set of solutions of $AB\xi = 0$ will then contain $m - \text{rk}(AB) - (m - \text{rk}(B)) = \text{rk}(B) - \text{rk}(AB)$ vectors ξ which are linearly independent and which are not solutions of $B\xi = 0$. Make up a matrix X of rank $\text{rk}(B) - \text{rk}(AB)$ with these vectors as columns. It satisfies the equation $A(BX) = 0$. Now use exercise 6 with B replaced by BX, and also use the right-hand member of the inequality.)

Note: The vector β is said to be a linear combination of the vectors $\alpha_1, \alpha_2, \dots, \alpha_m$, if numbers $c_1, \dots, c_m$ exist such that $\beta = c_1\alpha_1 + c_2\alpha_2 + \dots + c_m\alpha_m$.

8. Show that if one vector is a linear combination of a set of other vectors, then the vectors all together are linear dependent.

9. Show that if the vectors $\alpha_1, \alpha_2, \dots, \alpha_m$ are linearly dependent, then some one of them can be expressed as a linear

combination of the preceding ones in the sequence.

10. Let $\alpha_1, \alpha_2, \dots, \alpha_m$ be any set of linearly independent n -vectors. Prove that any other n -vector β can be expressed as a linear combination of $\alpha_1, \dots, \alpha_m$, and that the coefficients in this linear combination are uniquely determined.

(Hint: Treat the equation $\beta = c_1 \alpha_1 + c_2 \alpha_2 + \dots + c_m \alpha_m$ as a system of simultaneous linear equations in which the c 's are the unknowns. Use Exercise 4.)

Note: A vector space, or linear space, over the real or complex field is a set $\mathcal{G}$ of vectors, all with the same number of components, such that the sum of any two vectors always lies in $\mathcal{G}$, and such that if all the components of any one of the vectors are multiplied by any number in the field, the resulting vector still lies in $\mathcal{G}$. This means that all linear combinations of the given set of vectors lie in $\mathcal{G}$. The set of all n -vectors with components in a given field clearly forms a vector space over that field, so any subset of the set of all vectors which itself forms a vector space is often said to constitute a subspace rather than a space.

A vector space is said to be spanned by vectors $\alpha_1, \alpha_2, \dots, \alpha_m$ if these vectors themselves lie in the space, and if every other vector in it can be expressed as a linear combination of $\alpha_1, \alpha_2, \dots, \alpha_m$. We have seen above in Exercise 10 that any set of n linearly independent n -vectors spans the space of all such vectors. A set of linearly independent vectors which spans a vector space (or subspace) is called a basis for the space. As with the space of all vectors with a given number of components, any vector subspace generally has an in-

finite number of bases.

We now develop these concepts further by continuing the list of exercises.

11. Let V_n denote a certain vector space (the subscript n designates the number of components of each vector in the space), and let $\alpha_1, \dots, \alpha_m$ be a basis for V_n . Show that the coefficients of the representation of any vector β as a linear combination of $\alpha_1, \dots, \alpha_m$ (these are called the coordinates of the resolution of β in this basis) are uniquely determined. (Hint: Subtract two allegedly different resolutions and use the linear independence of the basis.)

12. Show that the set of unit n -vectors $e_1, e_2, \dots, e_n$ form a basis for the space of all real or complex n -vectors, and that the coefficients of the resolution of any vector in this basis are merely the components of the vector. (Note: In the case of complex vectors, it is to be understood that the components of e_1 are the complex numbers $1 + i0, 0 + i0, \dots, 0 + i0$, where $i = \sqrt{-1}$; and similarly for $e_2, \dots, e_n$.)

13. Prove that if vectors $\beta_1, \beta_2, \dots, \beta_k$ are each linear combinations of the vectors $\alpha_1, \alpha_2, \dots, \alpha_m$, $k > m$, then $\beta_1, \beta_2, \dots, \beta_k$ are linearly dependent. (Hint: The truth of the proposition becomes clear when the equations $\beta_i = c^{i1}\alpha_1 + c^{i2}\alpha_2 + \dots + c^{im}\alpha_m$, $i = 1, \dots, k$, are rewritten in matrix form $B = CA$, where B has the rows $\beta_1, \dots, \beta_k$, $C = [c^{ij}]$, and A has the rows $\alpha_1, \dots, \alpha_m$. Note that C is a $k \times m$ matrix, so its rank cannot exceed m . For a similar reason, that of A cannot exceed m . Now apply Section 1.3, Exercise 10, and then Theorem 1.18 .)

14. Prove that all of the bases for a given vector space V_n contain the same number of vectors. (Note: This number is therefore a characteristic of V_n , and is called the dimension of V_n . Hint: Let $\alpha_1, \dots, \alpha_p$ and $\beta_1, \dots, \beta_q$ be two different bases. First resolve the α 's in terms of the β 's and use Exercise 13; then resolve the β 's in terms of the α 's and use Exercise 13 again.)

15. Prove that there is always at least one basis for a given vector space V_n . (Hint: If V_n contains at least one non-zero vector, then since a single non-zero vector constitutes trivially a linearly independent subset of vectors, we can say that there is at least one subset of linearly independent vectors in V_n . Now V_n cannot contain a subset consisting of more than n linearly independent vectors. So there is some largest integer r , $1 \leq r \leq n$, such that V_n contains at least one set of r linearly independent vectors. Denote these vectors by $\alpha_1, \dots, \alpha_r$. Now show that the maximal character of r implies that any other vector in V_n must be a linear combination of $\alpha_1, \dots, \alpha_r$. Note: The reader may feel vaguely uneasy about this line of reasoning. If so, he can refer to Birkhoff-MacLane (1953), pp. 166-168 for another approach.)

^c1.6 Determinants. In this section, all matrices will be square. We shall define a certain function of a square matrix known as the determinant, and develop its principal properties.

The definition requires a prior understanding of what is meant by an inversion in a permutation of integers. Consider a permutation of positive integers, such as 5, 2, 3, 4, 1. An inversion is said to occur each time a larger integer precedes a

smaller one. Thus in 5, 2, 3, 4, 1 there are 7 inversions, counted as follows: (5 before 2) + (5 before 3) + (5 before 4) + (5 before 1) + (2 before 1) + (3 before 1) + (4 before 1).

The number of inversions in any permutation of integers can always be counted up in this way, and it is uniquely determined. This statement holds true whether or not there are gaps between the integers when they are arranged in order of magnitude.

A permutation of integers is called even if the number of inversions is even; otherwise it is called odd.

We now define the determinant of an $n \times n$ matrix $A = [a^{ij}]$. The determinant is the following polynomial in the elements a^{ij} :

$$(1.6.1) \quad \det(A) = \sum (\pm a^{i_1 1} a^{i_2 2} \dots a^{i_n n}),$$

where the summation is extended over all possible permutations $i_1, i_2, \dots, i_n$ of the row indices of the elements, and a + or - sign is used for a given term in the summation in accordance with whether the permutation of the row indices is even or odd. Each term of the summation in (1.6.1) contains one and only one factor from each row and each column of A.

When a matrix is displayed in full, or in a partitioned form, we shall replace the notation $\det ()$ by straight bars; thus:

$$\left| \begin{array}{c|c} A_1 & A_2 \\ \hline - & - \\ \hline A_3 & A_4 \end{array} \right|, \quad \left| \begin{array}{cc} a^{11} & a^{12} & a^{13} \\ a^{21} & a^{22} & a^{23} \\ a^{31} & a^{32} & a^{33} \end{array} \right|.$$

The reader may be familiar with these quick schemes for evaluating determinants in the cases $n = 2$ and $n = 3$:

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12}$$

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23} - a_{31}a_{22}a_{13} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33}$$

It will be noticed that the sign displayed before each term in the expansions is $+$ or $-$ according to whether the permutation of the row indices is even or odd.

The determinants of certain special types of square matrices with an arbitrary number of rows and columns can be evaluated directly from the definition of determinant, but in general cases it requires almost as much computation to evaluate a determinant as to solve a system of linear equations with the same number of rows. Two special types of matrices for which the value of the determinant can be read off from the definition are indicated below in the first two entries in a list of properties of determinants which we shall now develop.

Property 1. If A has a zero row or column, then $\det(A) = 0$.

The proof is obvious from inspection of (1.61) .

Property 2 . The determinant of a triangular matrix is equal to the product of the principal diagonal elements.

For let A be upper triangular. If we consider a typical term in (1.6.1), we can see at once that the only non-zero choice for a_{i1}^{11} is a^{11} ; but then the only permissible non-zero choice for a_{i2}^{12} is a^{22} , and so forth. The number of permutations in the row indices in the product $a^{11}a^{22}\dots a^{nn}$ is zero, which is even.

The remaining properties are concerned with the effects of elementary row and column operations on determinants

Property 3 . If each element of a row or column of A is multiplied by a scalar b , then the value of the determinant of the new matrix so obtained is $b \det(A)$.

The truth of this is again obvious from the definition.

It will be recalled that the operation of multiplying the i -th row of A by a scalar b is an elementary row operation of Type (ii), which can be accomplished by premultiplying A by a certain diagonal matrix D in which the elements of the principal diagonal are all ones except for the element in the i -th row, which is b . The value of $\det(D)$ is equal to b by Property 2, so Property 3 implies that $\det(DA) = \det(D)\det(A)$. Similarly it implies that $\det(AD) = \det(A)\det(D)$. A trivial application of mathematical induction now yields the following result:

Corollary . Let $D_1, D_2, \dots, D_k$ be a sequence of elementary matrices, each of which corresponds to a Type (ii) elementary operation, and let $E_1 = D_1D_2\dots D_k$. Then $\det(E_1A) = \det(D_1)\det(D_2)\dots\det(D_k)\det(A) = \det(E_1)\det(A) = \det(AE_1)$.

Property 4a . Interchanging two rows of a matrix changes the sign of its determinant.

Suppose first that two adjacent rows are interchanged, say the i -th and the $(i+1)$ -th . Consider a typical term in the summation in (1.6.1) . If the row index i happens to precede the index $i+1$ in this term, the term (without its sign) looks something like this:

$$a_{i_1}^{i_1} \dots a_{i_h}^{i_h} \dots a_{i+1,k}^{i+1,k} \dots a_n^{i_n} .$$

After the interchange of the i -th and $(i+1)$ -th rows, the same term looks like this :

$$a_{i_1}^{i_1} \dots a_{i+1,h}^{i+1,h} \dots a_{i,k}^{i,k} \dots a_n^{i_n} .$$

Clearly the number of inversions in the permutation of row indices was increased by exactly one by the row interchange. Similarly, if the index $i+1$ had originally preceded i , then the number of inversions in the permutation would have been decreased by one. In either case, the sign of the term must be changed, because if the permutation was odd before the row interchange, it was even afterwards, and vice versa. So the interchange of any two adjacent rows changes the sign of the determinant.

Now the interchange of any two rows whose row indices differ by, say, m , can be accomplished by $2m-1$ interchanges of two adjacent rows (m interchanges to bring one of the rows into the original position of the other row, and then $m-1$ interchanges to get the latter row back to where the first one was). Therefore the operation of interchanging these two rows changes the sign of the determinant $2m-1$ times, which

is the same as multiplying the determinant by

$$\overbrace{(-1)(-1) \dots (-1)}^{2m-1 \text{ factors}} = (-1)^{2m-1} = -1. \text{ This completes the proof.}$$

The operation of interchanging two rows of a matrix A is an elementary row operation of Type (i). It can be accomplished by premultiplying A by a certain special type of matrix P described in Section 1.3. To evaluate $\det(P)$, we specialize Property 4a to the case in which $A = I$. We find that $\det(P) = \det(PI) = -\det(I) = -1$. Then returning to the more general case, we have $\det(PA) = -1 \det(A) = \det(P)\det(A)$. From this, we obtain by a trivial induction the following result:

Corollary. Let $P_1, P_2, \dots, P_k$ denote matrices which correspond individually to Type (i) elementary row operations. Let $E_1 = P_1 P_2 \dots P_k$. Then $\det(E_1 A) = \det(P_1)\det(P_2) \dots \det(P_k)\det(A) = \det(E_1)\det(A)$.

Property 5a. If a matrix A has two identical rows, then $\det(A) = 0$.

Let B be the matrix obtained by interchanging the two identical rows. Clearly $B = A$, so $\det(B) = \det(A)$. But Property 4a states that $\det(B) = -\det(A)$. These equations are compatible only if $\det(A) = 0$.

Property 6 .

$$\begin{array}{cccc|cccc}
 a^{11} & a^{12} & \dots & a^{1n} & a^{11} & a^{12} & \dots & a^{1n} \\
 a^{21} & a^{22} & \dots & a^{2n} & a^{21} & a^{22} & \dots & a^{2n} \\
 \cdot & & & & \cdot & & & \\
 \cdot & & & & \cdot & & & \\
 \cdot & & & & \cdot & & & \\
 a^{i1} & a^{i2} & \dots & a^{in} & a^{i1} & a^{i2} & \dots & a^{in} \\
 \cdot & & & & \cdot & & & \\
 \cdot & & & & \cdot & & & \\
 \cdot & & & & \cdot & & & \\
 a^{n1} & a^{n2} & \dots & a^{nn} & a^{n1} & a^{n2} & \dots & a^{nn}
 \end{array}$$

$$+ \quad c \quad \begin{array}{cccc}
 a^{11} & a^{12} & \dots & a^{1n} \\
 a^{21} & a^{22} & \dots & a^{2n} \\
 \cdot & & & \\
 \cdot & & & \\
 \cdot & & & \\
 b^{i1} & b^{i2} & \dots & b^{in} \\
 \cdot & & & \\
 \cdot & & & \\
 \cdot & & & \\
 a^{n1} & a^{n2} & \dots & a^{nn}
 \end{array}$$

There is a similar property for the case in which a column instead of a row is expressed as the sum of two vectors .

The property is obvious from (1.6.1) .

Property 7a . If a row of the matrix A is multiplied by a number c and added to another row, the value of $\det(A)$ remains unchanged.

The proof consists in referring first to Property 6, then to Property 5a .

The operation described in Property 7a is an elementary row operation of Type (iii) . Such an operation is accomplished by premultiplying A by a certain unit triangle of the form $I + cU_{ij}$, where U_{ij} is a matrix consisting entirely of zeros except for a single element, which is a one. Since by Property 2, $\det(I + cU_{ij}) = 1$, it follows that $\det [(I + cU_{ij})A] = \det(A) = \det(I + cD_{ij})\det(A)$. Once again we have a corollary:

Corollary . Let E_3 be the product of a set of elementary matrices, each of which corresponds to a Type (iii) elementary row operation. Then $\det(E_3 A) = \det(A) = \det(E_3)\det(A)$.

We are now in a position to prove some fundamental theorems.

Theorem 1.20 . A necessary and sufficient condition for A to be non-singular is that $\det(A) \neq 0$.

For by Theorem 1.5a, there exists a non-singular matrix E_3 corresponding to a sequence of elementary row operations of Type (iii), such that $E_3 A = R$, where R is a semi-row-reduced matrix. By the corollary of Property 7a, $\det(E_3 A) = \det(A) = \det(R)$. If A is non-singular, then R has no zero rows, and it has one and only one non-zero element in each row and column. By the definition of determinant, the numerical value of $\det(R)$ is equal to the product of these non-zero elements. Therefore, in this case, $\det(R) \neq 0$. But if A is singular, then R has at

least one zero row, so by Property 1, $\det(R) = 0$. This completes the proof.

The leading elements of R were called the "pivots" of the construction used in the proof of Theorem 1.5a . Using this terminology, we can restate Theorem 1.20 in a form which is more useful for computational purposes.

Theorem 1.21 . If the $n \times n$ matrix A is singular, then $\det(A) = 0$. If A is non-singular, then $\det(A) = \pm a_{i_1 j_1} a_{i_2 j_2} \dots a_{i_n j_n}$, where $a_{i_1 j_1}, \dots, a_{i_n j_n}$ are the set of pivots used in any reduction of A to a semi-row-reduced form, and the $\pm$ or $-$ sign is chosen according to whether the number of inversions in the row indices is even or odd when the column indices of the pivots are arranged in ascending order.

To apply the theorem, it suffices to reduce A merely to a row-permuted echelon form. This is achieved by performing the steps indicated in the proof of Theorem 1.5a, but with this simplification: Once that multiples of a given row have been used in the construction, this row is left untouched in the subsequent steps. The reason that this simplification is possible when the objective is to evaluate $\det(A)$, is that the leading elements of the row-permuted echelon form thereby reached, are precisely the pivots which are used in a corresponding reduction of A to a semi-row-reduced form. Thus the numerical value of $\det(A)$ is equal to the product of the leading elements of any row-permuted echelon matrix to which A is row-equivalent. The sign of $\det(A)$ in the non-singular case is again determined in the manner described in Theorem 1.21 .

An alternative method of determining the sign to apply to the product of the pivots in the non-singular case consists in counting the number of interchanges of rows in the row-permuted echelon matrix necessary to achieve true echelon form. Property 4a will then give the sign.

The reduction of a matrix to row-permuted echelon form is one of the most efficient methods for the evaluation of determinants of large matrices.

Theorem 1.22 . The rank of any rectangular matrix is equal to the number of rows in the minor of maximal size having a non-vanishing determinant.

This is an immediate consequence of Theorems 1.11 and 1.20 . The size of the largest non-singular minor in a rectangular matrix --- that is, the size of the largest minor with a non-vanishing determinant --- is sometimes called the determinant rank of the matrix.*

In Table 1.1 we list the various possible definitions of the rank of a matrix for which the theoretical bases have been developed in this chapter. For real and complex matrices, the definitions all lead to the same number.

*
See for instance Birkhoff-MacLane (1953), p. 306 .

Table 1.1 Numerically equivalent definitions of the rank of a real or complex matrix A.

Special terminology	Definition	Reference
	The number of non-zero rows in any semi-row-reduced matrix to which A is row-equivalent	Section 1.3
Determinant rank	The size of the non-singular minor of maximal size	Theorem 1.11
Determinant rank	The size of the minor of maximal size having a non-vanishing determinant	Theorem 1.22
Row rank	The number of linearly independent rows of A	Theorem 1.18
Column rank	The number of linearly independent columns of A.	Theorem 1.22

Theorem 1.23 . The determinant of the product of a finite number of matrices is equal to the product of the determinants of the factors.

We shall show that $\det(AB) = \det(A) \det(B)$; thereafter the statement in the theorem follows immediately.

The theorem is trivial if either A or B is singular, for then by Exercise 10 of Section 1.3, AB is also singular, so by Theorem 1.20 its determinant is zero.

If neither A nor B is singular, then in accordance with Theorem 1.8 we express A as a product $A_3 A_1 A_2$, in which A_3 is

itself a product of elementary matrices corresponding to Type (iii) elementary row operations, A_1 is a product of elementary matrices corresponding to Type (i) elementary row operations, and A_2 is a product of elementary matrices corresponding to Type (ii) elementary row operations. Then $AB = A_3 A_1 A_2 B$. According to the Corollary of Property 7a, $\det(A_3 A_1 A_2 B) = \det(A_3) \det(A_1 A_2 B)$. According to the Corollary of Property 4a, $\det(A_1 A_2 B) = \det(A_1) \det(A_2 B)$. According to the Corollary of Property 3, $\det(A_2 B) = \det(A_2) \det(B)$. Thus $\det(AB) = \det(A_3) \det(A_1) \det(A_2) \det(B)$.

But with $B = I$, this states that $\det(A) = \det(AI) = \det(A_3) \det(A_1) \det(A_2) \cdot 1$. Therefore $\det(AB) = \det(A) \det(B)$, as was to be proved.

Corollary 1. Whether or not A and B are commutative, $\det(AB) = \det(BA)$.

Corollary 2. If A is non-singular, $\det(A^{-1}) = 1 / \det(A)$.

These two corollaries are obvious consequences of the theorem.

Corollary 3. $\det(A) = \det(A')$.

If A is singular, this statement is trivial, because by Exercise 9 of Section 1.3, A' is also singular, and by Theorem 1.20 $\det(A) = 0 = \det(A')$.

If A is non-singular, we express it as a product $A_3 A_1 A_2$ as in the proof of Theorem 1.23. Then $\det(A) = \det(A_3) \det(A_1) \det(A_2)$, and $\det(A') = \det(A_2' A_1' A_3') = \det(A_2') \det(A_1') \det(A_3')$. Now A_2 is a product of diagonal matrices, and so is itself a diagonal matrix. Therefore $A_2' = A_2$, and $\det(A_2') = \det(A_2)$.

The transpose of each factor in A_1 is again a matrix which performs a Type(1) elementary row operation; and the determinant of each such factor is -1 . There are just as many of these factors in the product A_1' as in A_1 , so $\det(A_1') = \det(A_1)$. Finally, A_3' , like A_3 , is a product of unit triangles, so $\det(A_3') = \det(A_3) = 1$. Putting all this together, we see that $\det(A') = \det(A_2') \det(A_1') \det(A_3') = \det(A_2) \det(A_1) \det(A_3) = \det(A)$, as was to be proved.

Elementary row operations on A' are elementary column operations on A , so by considering Properties 4a, 5a, and 7a in connection with Corollary 3 above, we obtain the following results:

Property 4b. (Same as Property 4a with the word "rows" replaced by "columns".)

Property 5b. (Same as Property 5a, with the word "rows" replaced by "columns".)

Property 7b. (Same as Property 7a, with the word "row" replaced by "column".)

Exercises

1. Prove that if a matrix has at least one non-zero element, its rank is at least equal to unity.
2. Show that if ξ and η are any two non-zero vectors, then the rank of the matrix $\xi\eta$ is one. (Hint: Use Exercise 1, and also Section 1.3, Exercise 10.)
3. Let A_1 be any $m \times m$ matrix and let A_2 be $(n-m) \times m$, $n > m$. Construct an $n \times n$ matrix A as follows:

$$A = \left[\begin{array}{c|c} A_1 & 0 \\ \hline A_2 & I_{n-m} \end{array} \right].$$

Prove that $\det(A) = \det(A_1)$. (Hint: Go back to the definition of determinant.)

4. Prove that

$$\begin{vmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{vmatrix} = \det(A_{11}) \det(A_{22}) .$$

(Hint: Verify that

$$\begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} = \begin{bmatrix} I & A_{12} \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} A_{11} & 0 \\ 0 & I \end{bmatrix}$$

$$= \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} \begin{bmatrix} A_{22} & 0 \\ A_{12} & I \end{bmatrix} \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} \begin{bmatrix} A_{11} & 0 \\ 0 & I \end{bmatrix} .)$$

5. Prove that if $A_{11}, A_{22}, \dots, A_{kk}$ are square matrices, then

$$\begin{vmatrix} A_{11} & A_{12} & \dots & A_{1k} \\ 0 & A_{22} & \dots & A_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_{kk} \end{vmatrix} = \det(A_{11}) \det(A_{22}) \dots \det(A_{kk}) .$$

6. Show that $\det(A^N) = (\det(A))^N$ for any positive integer N or 0 .

7. Prove that if A is a complex matrix, then $\det(\bar{A}) = \overline{\det(A)}$.

(Hint: Use (1.6.1) .)

8. Prove that the determinant of a hermitian matrix must be a real number. (Hint: Use Exercise 7.)
9. Prove that the determinant of a matrix is invariant under a similarity transformation. (For the definition of a similarity transformation, see the Note after Exercise 22 of Section 1.2 .)
10. Show that the determinant of a skew-symmetric matrix (see Exercise 12 of Section 1.2) is zero if the matrix has an odd number of rows. (Hint: Use Property 3 and Theorem 1.23 .)
11. Show that

$$\begin{vmatrix}
 1 & 1 & \dots & 1 \\
 a_1 & a_2 & \dots & a_n \\
 a_1^2 & a_2^2 & \dots & a_n^2 \\
 \vdots & \vdots & & \vdots \\
 a_1^{n-1} & a_2^{n-1} & \dots & a_n^{n-1}
 \end{vmatrix} = \prod_{j>i} (a_j - a_i) .$$

(Here for once the superscripts are exponents, not indices.)

Note: The determinant above is called Vandermonde's determinant, or an alternant. (Hint: The determinant must be a homogeneous polynomial of degree n in $a_1, a_2, \dots, a_n$. By Property 5b, it vanishes when $a_i = a_j$, $i, j = 1, \dots, n$. Therefore by the Remainder Theorem it has at least the factors indicated on the right side of the equation above, and perhaps other factors too. It remains to show that there cannot be any other factors.)

12. Consider the rectangular matrix

$$A = \begin{bmatrix} a & b & c & 0 & \dots & 0 & 0 \\ 0 & a & b & 0 & \dots & 0 & 0 \\ 0 & 0 & a & b & \dots & 0 & 0 \\ 0 & 0 & 0 & a & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & a & b \end{bmatrix}$$

$\overbrace{\hspace{10em}}^n$

$\underbrace{\hspace{10em}}_{n-1}$

Denote by A_k the minor obtained by omitting the k -th column of A . Prove that $d(A_k) = a^{k-1} b^{n-k}$. (Hint: Try partitioning A_k with dividing lines placed after the $(k-1)$ th row and column of A .)

Note. The following exercises provide bounds for determinants which are sometimes useful.

13.* Let $A = a^{ij}$ be a complex matrix which has a dominant diagonal**. Prove that

$$(1.6.2) \quad 0 < m_1 m_2 \dots m_n \leq |\det(A)| \leq \hat{m}_1 \hat{m}_2 \dots \hat{m}_n,$$

where $m_i =$

$$|a^{ii}| - \sum_{j=i+1}^n |a^{ij}|, \quad \hat{m}_i = |a^{ii}| + \sum_{j=i+1}^n |a^{ij}|, \quad i = 1, \dots, n.$$

Show also that an equality holds in (1.6.2) if and only if A is a lower triangular matrix. (Hint: The proof rests on a simple specialization of Exercise 10 of Section 1.4, which can be written as follows: If A is an $n \times n$ matrix with a dominant

* Price (1951). An improvement of (1.6.2) was given by Ostrowski (1952).

** This term is defined in Section 1.4, Exercise 8.

diagonal, then the system of $n-1$ equations in $n-1$ unknowns

$$(1.6.3) \sum_{j=2}^n a^{ij} x_j = -a^{i1}, \quad i = 2, 3, \dots, n,$$

has a unique solution $(x^2, \dots, x^n)$ such that $|x^i| < 1$, $i = 2, \dots, n$. [This is the special case of Exercise 10, Section 1.4, in which $y^i = -a^{i1}$, $r = 1$, and the number of equations displayed in that exercise is reduced by one.] For present purposes we

further identify the solution of (1.6.3) by subscripts in this manner: $(x_1^2, x_1^3, \dots, x_1^n)$. Now multiply the second column of A

by x_1^2 and add it to the first; also multiply the third column of A by x_1^3 and add it to the first; and so on to the last column, which is multiplied by x_1^n and added to the first. Because of

(1.6.3), all the elements of the first column are now zeros except the first element, which is $a^{11} + a^{12} x_1^2 + a^{13} x_1^3 + \dots +$

$a^{1n} x_1^n$; but by Property 7b the value of $\det(A)$ has not been changed. By Exercise 4, this value is

$$(1.6.4) \quad (a^{11} + a^{12} x_1^2 + \dots + a^{1n} x_1^n) \begin{vmatrix} a^{22} & \dots & a^{2n} \\ \vdots & & \vdots \\ a^{n2} & \dots & a^{nn} \end{vmatrix}.$$

But the matrix of the determinant on the right in (1.6.4) satisfies the same hypotheses as A , so the process can be repeated until we arrive at this formula:

$$(1.6.5) \quad \det(A) = (a^{11} + a^{12} x_1^2 + \dots + a^{1n} x_1^n) (a^{22} + a^{23} x_2^3 + \dots + a^{2n} x_2^n) \dots a^{nn}.$$

All of the numbers x_k^i appearing here satisfy the inequality

$|x_k^i| < 1$. The reader should have no difficulty in completing the proof from this point.

14.* Let A be a real matrix with a dominant diagonal and with positive principal diagonal elements. Then $\det(A) > 0$ and (1.6.2) is valid with the absolute value indication on $\det(A)$ removed. (Hint: Use (1.6.5) .)

15.** Let A be a Hermitian matrix with a dominant diagonal and with positive principal diagonal elements. Then $\det(A) > 0$ and (1.6.2) is valid with the absolute value indication on $\det(A)$ removed. (Hint: We know from Exercise 8. that $\det(A)$ is real. Since this is so, and since for the same reason the determinant on the right in (1.6.4) is real, it follows that the first factor in (1.6.4) is real. By repeating the argument, it can be shown that all the factors in (1.6.5) are real. The remainder of the proof is left to the reader.

1.7. Cramer's Rule. Throughout this section $A = [a^{ij}]$ will denote a real or complex $n \times n$ matrix.

There is a famous formula, known as Cramer's Rule, which exhibits each component of the solution of a system $A\xi = \eta$ of linear equations with a non-singular matrix as a linear combination of the components of η . We shall now develop this formula.

Let A_{ij} be the $(n-1)$ -rowed minor of A obtained by deleting the i -th row and j -th column of A . Let $b^{ij} = (-1)^{i+j} \det(A_{ij})$; this is called the cofactor of the element a^{ij} .

Consider now the definition of $\det(A)$ as given by (1.6.1) .
For convenience, we repeat the formula here:

$$(1.7.1) \quad \det(A) = \sum (\pm a^{i_1 1} a^{i_2 2} \dots a^{i_n n}) .$$

What is the value of the sum of all the terms on the right side which contain $a^{1,1}$? Obviously the sign of any such term will be determined by the number of inversions in the indices $i_2, i_3, \dots, i_n$, and the terms themselves are precisely the terms in the expansion of $\det(A_{i_j})$. Thus this partial sum equals $a^{11} \det(A_{i_j}) = a^{11} c^{11}$.

Now what is the sum of all the terms in (1.7.1) which contain the element a^{hk} ? We can answer this easily by using Properties 4a and 4b. We first interchange the h -th row of A with the $(h-1)$ th, then the $(h-1)$ th with the $(h-2)$ th, and so forth, until with $h-1$ interchanges of adjacent rows we get the h -th row in the first position. Then we similarly pass the k -th column into the first position. The effect of all this is to make a^{hk} become the element in the first row and column of a new matrix B in which all the rows and columns are in the same order as in A except that the h -th row and k -th column of A are the first row and first column of B . In the expansion of B , the sum of all the terms containing the element in the first row and column, is equal to $a^{hk} \det(A_{hk})$. Moreover, by Properties 4a and 4b, $\det(A) = (-1)^{h-1 + k-1} \det(B) = (-1)^{h+k} \det(B)$. Therefore the sums of all the terms in the expansion of A containing the element a^{hk} is $a^{hk} (-1)^{h+k} \det(A_{hk}) = a^{hk} c^{hk}$.

But now if we first add up all the terms in (1.7.1) which contain the first element of a given row or of a given column, and then add up all the terms which contain the second element

of the given row or the given column, and so on through all the elements of the given row or column, then these partial sums all together include all the terms in the summation in (1.7.1) . But these partial sums have no terms in common. Therefore the sum of the partial sums must be exactly equal to $\det(A)$.

Thus we have proved the following result:

Theorem 1.24 . For any integers h and k , $1 \leq h \leq n$, $1 \leq k \leq n$,
 $\det(A) = a_{h1}^{h1} c_{h1}^{h1} + a_{h2}^{h2} c_{h2}^{h2} + \dots + a_{hn}^{hn} c_{hn}^{hn} = a_{1k}^{1k} c_{1k}^{1k} + a_{2k}^{2k} c_{2k}^{2k} + \dots + a_{nk}^{nk} c_{nk}^{nk}$.

The two sums appearing in the theorem are called respectively the expansion of $\det(A)$ according to cofactors of the h -th row and the expansion of $\det(A)$ according to cofactors of the k -th column.

Corollary: $\frac{\partial \det(A)}{\partial a^{ij}} = c^{ij}$, $i, j = 1, \dots, n$.

The truth of this is obvious from Theorem 1.24, and from the fact that all the cofactors used in the expansion of $\det(A)$ according to cofactors of the i -th row are independent of a^{ij} .

The corollary suggests that if the cofactor of a particular element a^{ij} of a matrix A is relatively large, then a small change in a^{ij} might mean a large change in $\det(A)$. The following example, due to John Todd, illustrates this point:

$$\begin{vmatrix} -73 & 78 & 24 \\ 92 & 66 & 25 \\ -80 & 37 & 10 \end{vmatrix} = 1 ; \quad \begin{vmatrix} -73 & 78 & 24 \\ 92 & 66 & 25 \\ -80 & 37 & 10.01 \end{vmatrix} = -118.94$$

If in the matrix A the i -th row is replaced by the h -th row, but the h -th row is left untouched, then we obtain a new matrix B

such that by Property 5a, $\det(B) = 0$. In B , the cofactors of the elements a^{hj} , $j = 1, \dots, n$, in the i -th row are of course identical with the cofactors c^{ij} of the elements of the i -th row in A .) The matrices A and B differ only in their i -th rows, and it will be recalled that the i -th row is crossed out in setting up the cofactors of the elements of the i -th row.) Expanding B according to the cofactors of its i -th row, we obtain the first equation in the following theorem. The second one is proved by a similar argument.

Theorem 1.25 . For any integers h and i , $n \neq i$, $1 \leq h \leq n$, $1 \leq i \leq n$, $a^{h1}c^{i1} + a^{h2}c^{i2} + \dots + a^{hn}c^{in} = 0$, and $a^{1h}c^{i1} + a^{2h}c^{i2} + \dots + a^{nh}c^{in} = 0$.

The two sums in Theorem 1.25 are called expansions of $\det(A)$ by alien cofactors.

The $n \times n$ matrix $[c^{ij}]'$ is called the adjoint of A . We shall denote it by $\text{adj } A$. (The reader should note carefully that $\text{adj } A$ is not $[c^{ij}]$, but rather, the transpose of $[c^{ij}]$.) The adjoint has the following important property:

Theorem 1.26 . $A(\text{adj } A) = (\text{adj } A)A = \det(A) \cdot I$. If A is non-singular, $A^{-1} = (1/\det(A))\text{adj } A$.

The proof consists merely in multiplying out the matrix products $A[c^{ij}]'$ and $[c^{ij}]'A$ and using Theorems 1.24 and 1.25 .

We now come to Cramer's Rule:

Theorem 1.27 . The i -th component x^i of the solution ξ of the system of linear equations $A\xi = \eta$, where $A = [a^{ij}]$ is non-singular, is given by either of the formulas:

$$(1.7.2) \quad x^i = \frac{1}{\det(A)} (c^{1i}y^1 + c^{2i}y^2 + \dots + c^{ni}y^n),$$

$$(1.7.3) \quad x^i = \frac{1}{\det(A)} \begin{vmatrix} a^{11} & a^{12} & \dots & a^{1,i-1} & y^1 & a^{1,i+1} & \dots & a^{1n} \\ a^{21} & a^{22} & \dots & a^{2,i-1} & y^2 & a^{2,i+1} & \dots & a^{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ a^{n1} & a^{n2} & \dots & a^{n,i-1} & y^n & a^{n,i+1} & \dots & a^{nn} \end{vmatrix},$$

$i = 1, \dots, n$, where $\eta = (y^1, \dots, y^n)$ and c^{ij} is the cofactor of a^{ij} .

The proof of (1.7.2) follows from the second statement in Theorem 1.26. For we have, $\xi = A^{-1}\eta = (1/\det(A))[c^{ij}]'\eta$, which indeed may be taken as a compact summary of Cramer's Rule. The equivalence of (1.7.3) with (1.7.2) is established by expanding the determinant in the numerator on the right of (1.7.3) according to cofactors of the i -th column.

Cramer's Rule furnishes a convenient method for solving systems of simultaneous linear equations consisting of no more than three or four equations. But except in special cases Cramer's Rule requires much more computation than certain other methods if the number of equations is higher than four, even when the most efficient methods of evaluating determinants are used. (See Table 2.2.) Therefore we are not including it in our formal list of practical methods for the solution of simultaneous linear equations, which begins in Chapter 2.

Corollary. If x^i is the i -th element of the solution of $A\xi = \eta$,

then $\frac{\partial x^i}{\partial y^j} = \frac{c^{ji}}{\det(A)} = b^{ij}$, where $\eta = (y^1, \dots, y^n)$ and

$$[b^{ij}] = A^{-1}.$$

The proposition is an obvious consequence of (1.7.2).

1. Prove that if $\det(A) = 0$, then the vectors $\xi_k = (c^{k1}, c^{k2}, \dots, c^{kn})$, $k = 1, 2, \dots, n$, are all solutions of the problem $A \xi = 0$.
2. Prove that if A is non-singular, then $\det(\text{adj } A) = [\det(A)]^{n-1}$.
3. Prove that a necessary and sufficient condition for $\text{adj } A$ to be singular is that A is singular. (Hint: The necessity follows at once from Theorems 1.23 and 1.26. In proving the sufficiency, the hypothesis states that $(\text{adj } A)A = 0$, where 0 stands for an $n \times n$ zero matrix. Assume that $\text{adj } A$ is non-singular and use Section 1.3, Exercise 3, and Section 1.6, Exercise 1, to arrive at a contradiction.)
4. Show that

$$\begin{vmatrix} -x & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & -x & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & -x & 1 & \dots & 0 & 0 \\ 0 & 0 & 0 & -x & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -x & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & -a_{n-3} & \dots & -a_2 & -(a_1 + x) \end{vmatrix} =$$

$(-1)^n (x^n + a_1 x^{n-1} + a_2 x^{n-2} + \dots + a_n)$. (Hint: Expand by

cofactors of the last row, starting at the right. Use Exercise 12 of Section 1.6 to evaluate the cofactors.)

5. Show that $\det [\text{adj}(\text{adj } A)] = [\det(A)]^{(n-1)^2}$.

1.8 . Problems involving complex matrices. Let A be a rectangular complex matrix with elements $a_1^{hk} + ia_2^{hk}$, $h = 1, \dots, n$, $k = 1, \dots, m$, where a_1^{hk} and a_2^{hk} are real numbers and $i = \sqrt{-1}$. We can write A as $A_1 + iA_2$, where $A_1 = [a_1^{hk}]$ and $A_2 = [a_2^{hk}]$.

Note that $\bar{A} = A_1 - iA_2$, $A' = A_1' + iA_2'$.

The algebra of the special 2×2 matrices of the form

$$\begin{bmatrix} a_1 & -ia_2 \\ a_2 & a_1 \end{bmatrix}$$

is a representation of the algebra of complex numbers $a_1 + ia_2$.

Similarly, the algebra of special $2n \times 2n$ partitioned matrices of the form

$$(1.8.1) \quad A^* = \begin{bmatrix} A_1 & -A_2 \\ A_2 & A_1 \end{bmatrix}$$

is a representation of the algebra of complex matrices $A = A_1 + iA_2$.

For example, if A is $n \times m$ and B is $m \times p$, then $AB = (A_1 + iA_2)(B_1 + iB_2)$

$= A_1B_1 - A_2B_2 + i(A_1B_2 + A_2B_1)$, and at the same time

$$\begin{bmatrix} A_1 & -A_2 \\ A_2 & A_1 \end{bmatrix} \begin{bmatrix} B_1 & -B_2 \\ B_2 & B_1 \end{bmatrix} = \begin{bmatrix} A_1B_1 - A_2B_2 & -(A_1B_2 + A_2B_1) \\ A_1B_2 + A_2B_1 & A_1B_1 - A_2B_2 \end{bmatrix}.$$

This suggests that many problems in linear computation

involving complex matrices might be performed just as well by replacing the complex matrices by partitioned real matrices of the form (1.8.1) . There are however a number of limitations on the usefulness of this idea. Some of them will become apparent when we are in a position to compare the relative efficiencies of various alternative computation programs aimed at solving the same problem. (An extended discussion of this matter in the complex case will be found below in Section 3.4 .) Other limitations may occur to the reader when he examines the rather complicated nature of some of the relationships which exist on the one hand between the inverse and the determinant of a complex matrix, and on the other, the inverse and determinant of the real and imaginary parts of the matrix. The remainder of this section will be devoted to developing these relationships.

Given the $n \times m$ complex matrix A , and the $m \times p$ complex matrix $Y = Y_1 + iY_2$, it can be seen by equating real and imaginary parts that the problem of finding an $m \times p$ matrix (or matrices) $X = X_1 + iX_2$ such that $AX = Y$, is exactly the same as that of finding a real matrix X^* of the form (1.8.1) such that $A^*X^* = Y^*$. In more detail, the equivalent problem is to find X_1 and X_2 in

$$(1.8.2) \quad \left[\begin{array}{c|c} A_1 & -A_2 \\ \hline A_2 & A_1 \end{array} \right] \left[\begin{array}{c|c} X_1 & X_2 \\ \hline X_2 & X_1 \end{array} \right] = \left[\begin{array}{c|c} Y_1 & -Y_2 \\ \hline Y_2 & Y_1 \end{array} \right],$$

where all matrices are real.

There is a certain amount of redundancy in the statement of the problem given by (1.8.2) . A more compact form is

$$(1.8.3) \quad \left[\begin{array}{c|c} A_1 & -A_2 \\ \hline A_2 & A_1 \end{array} \right] \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} .$$

Now let A be $n \times m$. By specializing Y to the matrix $Y = I_n \otimes i.0$, we obtain the first part of the following result, which is obvious from (1.8.3) and the definitions:

Theorem 1.28 . Alternative necessary and sufficient conditions for a complex matrix $A = A_1 \otimes iA_2$ to be non-singular are that

- (a) The matrix A^* is non-singular, where A^* is given by (1.8.1).
- (b) $\bar{A}$ is non-singular.
- (c) $A_1^2 \otimes A_2^2 \otimes (A_2 A_1 - A_1 A_2)$ is non-singular.

The proof of (b) is to be found in the fact that $\det(\bar{A}) = \overline{\det(A)}$. (See Exercise 7 of Section 1.6 .) As for (c), we have to note that $(A_1 \otimes iA_2)(A_1 - iA_2) = A_1^2 \otimes A_2^2 \otimes i(A_2 A_1 - A_1 A_2)$, so by Theorem 1.23, $\det(A) \det(\bar{A}) = \det A_1^2 \otimes A_2^2 \otimes i(A_2 A_1 - A_1 A_2)$.

The interest in Theorem 1.28 (c) lies chiefly in the case in which A_1 and A_2 are commutative, for then the imaginary part is zero. We return to this case in Exercise 2 below.

The availability of relatively simple formulas in terms of A_1 and A_2 for A^{-1} , A^{*-1} (when these inverses are known to exist), and for $\det(A^*)$ seems to hinge on the non-singularity of A_1 or A_2 . Note first that with A_1 non-singular, we have the identity

$$(1.8.4) \quad A^* = \left[\begin{array}{c|c} I_n & 0 \\ \hline A_2 A_1^{-1} & I_n \end{array} \right] \left[\begin{array}{c|c} A_1 & -A_2 \\ \hline 0 & A_1 + A_2 A_1^{-1} A_2 \end{array} \right].$$

Thus by Theorem 1.23 and Exercise 4 of Section 1.6,

$$(1.8.5) \quad \det(A^*) = 1 \cdot \det(A_1) \det(A_1 + A_2 A_1^{-1} A_2) = 1 \cdot \det(A_1) \cdot \det(A_1^{-1}) \det \left[I_n + (A_1^{-1} A_2)^2 \right] = \det \left[I_n + (A_1^{-1} A_2)^2 \right].$$

If A_2 is non singular, then by a similar argument we find that $\det(A^*) = \det \left[I_n + (A_2^{-1} A_1)^2 \right]$.

A formula for the inverse of A^* can be derived by methods which will be presented in Chapter 3. Here we shall merely write down the formula without motivation. The reader can of course verify that it is true by multiplication by A^* .

$$(1.8.6) \quad A^{*-1} = \left[\begin{array}{c|c} B^{-1} & A_1^{-1} A_2 B^{-1} \\ \hline -A_1^{-1} A_2 B^{-1} & B^{-1} \end{array} \right],$$

where $B = A_1 + A_2 A_1^{-1} A_2$. It follows that $A^{-1} = B^{-1} - i A_1^{-1} A_2 B^{-1}$.

There is no simple connection between the singularity or non-singularity of $A = A_1 + i A_2$ or A^* and that of A_1 and of A_2 . For instance, A can be non-singular with both A_1 and A_2 singular, as the trivial example

$$A = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + i \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

shows. (In such a case, formulas (1.8.5) and (1.8.6) are not

applicable.) Also, A can be singular with both A_1 and A_2 non-singular, as is demonstrated by

$$A = \begin{bmatrix} 1 + i & 1 \\ 2 & 1 - i \end{bmatrix}.$$

The reader will have no difficulty in constructing similarly negative examples for all of the other possible combinations.

Exercises .

1. Show that if $A_2 = cA_1$, where c is a real scalar, then A and A^* are singular or non-singular with A_1 , and the rank of A^* is twice the rank of A_1 .

2. Show that if A_1 and A_2 are commutative, then $|\det(A)|^2 =$

$\det(A^*) = \det(A_1^2 + A_2^2)$. Show by a counter-example that the proposition is not true if A_1 and A_2 are not commutative.

3. Show that each of the following is a necessary and sufficient condition for A to be hermitian:

(a) A_1 is symmetric, A_2 is skew-symmetric;

(b) A^* is symmetric.

(Note that if A has an odd number of rows and is hermitian, then A_2 must be singular; but this has no bearing on whether A and A^* are singular.)

(Dupl)

(Dupl)

220 Curtis
AUTHOR

TITLE Matrix inversion and
the solution of linear

INDEX

25 V.

N. B. L.

days.

144220 Date Due

OCT 29 '59

AUG 10 87

PRINTED IN U. S. A.

43

